DNA Methylation Signatures of Depressive Symptoms in Middle-aged and Elderly Persons

Title: DNA methylation signatures of depressive symptoms in middle-aged and elderly persons identified in a large multi-ethnic meta-analysis of epigenome-wide studies

Subtitle: *An epigenome-wide association study of depression*

Date revised: 17-05-2018

Olivera Jovanova[1†], PhD, Ivana Nedeljkovic[1†], MD, Spieler Derek[2, 3†], MD, Rosie M. Walker[4, 5†], PhD, Chunyu Liu[6, 7, 8†], PhD, Michelle Luciano[4, 9†], PhD, Jan Bressler[10], PhD, Jennifer Brody[11], BA, Amanda J. Drake[12], PhD, Kathryn L. Evans[4, 5], PhD, Rahul Gondalia[13], MPH, Sonja Kunze[2, 14], PhD, Brigitte Kuhnel[2, 14], MSc, Jari Lahti[15], PhD, Rozenn N. Lemaitre[11], PhD, Riccardo E. Marioni[4, 9], PhD, Brenton Swenson[11, 16], MSc, Jayandra Jung Himali[17], PhD, Hongsheng Wu[18], PhD, Yun Li[19, 20, 21], PhD, Allan F. McRae[22], PhD, Tom C. Russ[4, 23, 24], PhD, James Stewart[13, 25], MSc, Zhiying Wang[10], MSc, Guosheng Zhang[19, 20, 21], MSc, Karl-Heinz Ladwig[2, 3], PhD, Andre G. Uitterlinden[1, 26], PhD, Xiuqing Guo[27], PhD, Annette Peters[2, 14], PhD, Katri Räikkönen[15], PhD, John M. Starr[4, 23], PhD, Melanie Waldenberger[2, 14], PhD, Naomi R. Wray[22], PhD, Whitsel A. Eric[13, 28], MD, Nona Sotoodehnia[11], MD, Sudha Seshadri[6, 29], MD, David J. Porteous[4, 5], PhD, Joyce van Meurs[26], PhD, Thomas H. Mosley[30], PhD, Andrew M. McIntosh[4, 31], PhD, Michael M. Mendelson[6, 7, 32], MD, Daniel Levy[6, 7], MD, Lifang Hou[33], PhD, Johan G. Eriksson[34], PhD, Myriam Fornage[10, 35], PhD, Ian J. Deary[4, 9], PhD, Andrea Baccarelli[36], PhD, Henning Tiemeier[1, 37, 38*], PhD, Najaf Amin[1], PhD.

[1] Department of Epidemiology, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[2] Institute of Epidemiology II, Helmholtz Zentrum München, 85764 Neuherberg, Germany.
[3] Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie des Klinikums Rechts der Isar der TUM, 81675 Munich, Germany.
[4] Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK.
[5] Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, The University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK.
[6] The Framingham Heart Study, Framingham, MA, USA.

[7] The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA.
[8] Boston University School of Public Health, 715 Albany St, Boston, MA, USA.
[9] Psychology, University of Edinburgh, Edinburgh, UK.
[10] Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.
[11] Cardiovascular Health Research Unit, Department of Medicine, University of Washington.
[12] University/British Heart Foundation Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh.
[13] Department of Epidemiology, University of North Carolina. Chapel Hill, NC 27514.
[14] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany.
[15] Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland.
[16] Institute for Public Health Genetics, School of Public Health, University of Washington.
[17] Boston University School of Medicine and of Public Health, The Framingham Study, 72 East Concord Street, B 604, Boston, MA 021118, USA.
[18] Computer Science and Networking, Wentworth Institute of Technology, Boston, MA.
[19] Department of Genetics, University of North Carolina. Chapel Hill, NC 27514.
[20] Department of Biostatistics, University of North Carolina. Chapel Hill, NC 27514.
[21] Department of Computer Science, University of North Carolina. Chapel Hill, NC 27514.
[22] Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia; Queensland Brain Institute, The University of Queensland, Brisbane, Australia.
[23] Alzheimer Scotland Dementia Research Centre, 7 George Square, Edinburgh, UK. EH8 9JZ.
[24] Centre for Dementia Prevention, University of Edinburgh, UK.
[25] Carolina Population Center, University of North Carolina. Chapel Hill, NC 27514.
[26] Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3000 CA, The Netherlands.
[27] The Institute for Translational Genomics and Population Sciences, Department of Pediatrics at Harbor-UCLA Medical Center.
[28] Department of Medicine, University of North Carolina. Chapel Hill, NC 27514.
[29] Boston University School of Medicine,72 East Concord St, B602, Boston, MA 02118.
[30] MIND Center, University of Mississippi Medical Center, Jackson, MS.
[31] Division of Psychiatry, The University of Edinburgh, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, UK.
[32] Department of Cardiology, Boston Children's Hospital, Harvard Medical School, Boston, MA, US.
[33] Northwestern University. 680 N Lake Shore Drive, Suite 1400, Chicago IL 60611.
[34] Department of General Practice and Primary health Care, University of Helsinki, Tukholmankatu 8 B, Helsinki, 00014, Finland.
[35] Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX.
[36] Harvard University. 665 Huntington Ave, Boston, MA 02115.
[37] Department of Child and Adolescent Psychiatry, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[38] Department of Social and Behavioral Science, Harvard TH Chan School of Public Health, Boston, USA.
[†] Equal contributors.
[*] Corresponding author.

Corresponding author: Henning Tiemeier, MD, PhD,
Epidemiology, Erasmus MC-University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands, T:+31 107043489, F:+31 107044657.
Email: h.tiemeier@erasmusmc.nl

Word count: 3,457

Key Points

Question: Can we identify DNA methylation signatures of depressive symptoms in middle-aged and elderly persons in the general populations in blood?

Findings: In this large epigenome-wide association study of depressive symptoms, comprising of 11,256 participants of European and African origin from 11 population-based cohorts, we identified three genomic sites significantly associated with depressive symptoms. All three sites are either in genes or have downstream effects on genes expressed in the brain, and converge to the axon guidance pathway.

Meaning: Our study suggests that robust DNA methylation signatures of depression are identifiable in blood and these signatures may be similar across various ethnicities.

Abstract

Importance: Depressive disorders arise from a combination of genetic and environmental risk factors, however, the pathophysiology and underlying molecular events leading to depression remain elusive. Epigenetic disruption provides a plausible mechanism through which gene-environment interactions lead to depression. Large-scale epigenome-wide studies on depression are missing, hampering the identification of potentially modifiable biomarkers.

Objective: To identify robust epigenetic mechanisms underlying depression in middle-aged and elderly persons using DNA methylation in blood.

Design: We performed the first cross-ethnic meta-analysis of epigenome-wide association studies (EWAS) within the framework of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium.

Setting: Discovery EWAS was performed in nine population-based cohorts. Results of the EWAS from all cohorts were pooled using sample-size weighted meta-analysis. Replication of the top epigenetic sites from the discovery stage was performed in two independent population-based cohorts.

Participants: The discovery sample included 7,948 individuals of European origin and the replication sample included 3,308 individuals of African-American and European origin. Only participants that were assessed for both depressive symptoms and whole blood DNA methylation were included in the study.

Outcome: Whole blood DNA methylation levels were assayed with Illumina-Infinium Human Methylation 450K BeadChip and depressive symptoms were assessed by questionnaire.

Results: The discovery cohorts consisted of 7,948 individuals (48% female) with a mean age of 65.4 (SD=5.8) years. The replication cohort consisted of 3,308 individuals (74% female) with a mean age of 60.3 (SD=6.4) years. The EWAS identified methylation of three CpG sites including cg04987734 (*p-value*=1.57×10^{-8}, *n*=11256, *CDC42BPB* gene), cg12325605 (*p-value*=5.24×10^{-9}, *n*=11256, *ARHGEF3* gene) and an intergenic CpG site cg14023999 (*p-value*=5.99×10^{-8}, *n*=11256, chromosome=15q26.1) significantly associated with increased depressive symptoms. The predicted expression of *CDC42BPB* and *ARHGEF3* was significantly associated with major depression in brain and fibroblasts, respectively.

Conclusion: We report the first robustly associated methylated sites for depressive symptoms. All three findings point towards axon guidance as the common disrupted pathway in depression. Our findings provide new insights into the molecular mechanisms underlying the complex pathophysiology of depression. Further research is warranted to determine the utility of these findings as biomarkers of depression and evaluate any potential role in the pathophysiology of depression and their downstream clinical effects.

Introduction

Depression is one of the most common mental health disorders that is projected to play a leading role in disease burden by the year 2030[1]. In later life, depression is associated with disability, increased mortality, dementia and poor outcomes from physical illness[2]. Further, more people aged over 65 years commit suicide than in any other age group, and most have major depression[3]. Limited understanding of the molecular mechanisms underlying depression is a major bottleneck in the development of innovative treatment, prognostic markers, and prevention strategies.

Studying depression is challenging, as it is a heterogeneous disorder with a multifactorial etiology[4]. This heterogeneity increases with age as the incidence of chronic diseases and disability rises. The contribution of genetics to the risk of depression is moderate with heritability estimates ranging from 40 to 50%[5] and modest (18%) in the elderly[6]. Genome-wide association studies (GWAS) have recently identified numerous rare and common genetic variants associated with depression and related traits[7-10]. However, genetic variation alone does not completely explain an individual's risk for developing depression. Among environmental factors, adverse life-events and stress are major risk factors for depression[11]. Converging evidence from animal and human studies suggest that psychosocial stressors trigger depression onset by inducing elevations in pro-inflammatory cytokines[12]. These psychosocial stressors are also known to influence epigenetic mechanisms, such as DNA methylation[13] that can drive sustained changes in gene expression[14]. The high contribution of environmental factors to

depression in the elderly makes DNA methylation an interesting candidate mechanism for studies of the molecular basis of late-life depression.

DNA methylation may be global or tissue-specific[15]. Tissues likely to be involved in complex psychiatric disorders, such as brain, are not directly accessible from living patients. The, use of post-mortem brain tissue to study DNA methylation is a possible solution, although obtaining a sufficient sample size is challenging[16]. To study differential DNA methylation associated with mental health symptoms on a large scale, peripheral tissues such as blood constitutes a useful proxy for detecting trans-tissue changes and the most appropriate tissue for biomarkers[16,17]. Moderate correlation has been demonstrated between blood and brain tissues at non-tissue specific regulatory regions across the methylome[18]. To date, several studies have assessed the correlation between depression and blood DNA methylation[19,20]. However, these studies have been limited to a small number of DNA methylation sites (CpGs) and/or small samples. For instance, the largest published epigenome-wide association study (EWAS) assessed brain DNA methylation in 76 cases persons who died during a depressive episode and 45 controls[21]. Moreover, these studies compared depressed cases with healthy controls. Depression represents an arbitrarily selected extreme of the continuum of varying severity and duration[22], whereas a broad phenotype approach can be more representative for the general population. In a large study consisting of 252,503 individuals from 68 countries showed that sub-threshold depressive disorders produce significant decrements in health and do not qualitatively differ from full-blown episodes of depression[23]. A meta-analysis in individuals aged over 55 found two to three times higher prevalence of sub-threshold depressive symptomology than major

depression[24]. Use of rating scales have therefore been recommended for the assessment of depressive problems in the elderly[2].

In the current study, we performed EWAS of depressive symptoms using whole blood samples of 7,948 individuals of European ethnicity from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. We replicated our findings in 3,308 individuals of African-American and European ancestry. Finally, we used publicly available expression quantitative methylation (eQTM) loci and expression quantitative loci (eQTL) databases to identify the downstream effects of the associated methylation loci.

Materials and Methods

Study population

The study sample for the discovery analysis included a total of 7,948 participants of European ancestry from nine population-based cohorts of the CHARGE consortium (**Table 1**): Cardiovascular Health Study (CHS)[25], Framingham Heart Study (FHS)[26], Helsinki Birth Cohort Study (HBCS)[27], Cooperative Health Research in the Augsburg Region (KORA) study[28], two sub-cohorts from Lothian Birth-Cohort born in 1921 (LBC1921)[29] and 1936 (LBC1936)[30], two sub-cohorts from Rotterdam Study (RS-III and RS-BIOS)[31] and Generation Scotland: Scottish Family Health Study (GS) study[32]. These cohorts included community dwelling individuals, who were not selected based on disease status. Informed consent was obtained from all participants. The same cohorts have been successfully used to identify differentially methylated sites associated with cognitive traits[33], inflammation[34] and smoking[35]. The protocol for each study was approved by the institutional review board of each institution.

The replication sample included 3,308 participants, largely of African American origin from the Atherosclerosis Risk in Communities Study (ARIC)[36] and European origin from the Women's Health Initiative - Epigenetic Mechanisms of PM-Mediated Cardiovascular disease (WHI-EMPC) that joined the consortium later for the replication phase of the study[37]. Detailed information for each cohort is provided in the **Supplementary Text**.

Depressive symptoms assessment

Depressive symptoms were measured using self-reported questionnaires or structured interview performed by a trained researcher, psychologist, or psychiatrist at the same time point when blood samples were obtained for DNA methylation quantification (**Table 1**). Four cohorts (FHS, HBCS, RS-III, and RS-BIOS) assessed depressive symptoms using the 20-item Centre for Epidemiologic Studies Depression (CES-D) scale[38], while CHS used the 10-item CES-D scale. Participants could score from zero to 60 (or 30 for CHS) points, where higher scores suggest more depressive symptoms. WHI-EMPC used a cohort specific CES-D/DIS screening instrument, which is described in detail in the **Supplementary Text**. The LBC1921 and LBC1936 assessed self-reported depressive symptoms using the Hospital Anxiety and Depression Scale-depression subscale (HADS-D)[39], which consists of seven items. Participants could score from zero to 21. The KORA study used the self-administered Patient Health Questionnaire (PHQ-9)[40] representing a depression module that scores each of the nine Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) criteria for depression from zero to three. The GS study assessed life-time history of depression using the Structured Clinical Interview for DSM-IV Disorders (SCID)[41]. The ARIC study assessed depressive symptoms using the 21-item

Maastricht Questionnaire (21-MQ). In all cohorts depressive symptoms were analyzed as continuous variable except for GS, which studied depression status as binary trait.

DNA methylation sample and measurement

In all cohorts, DNA was extracted from whole blood and methylation levels were assessed using the Illumina-Infinium Human Methylation 450K BeadChip (Illumina Inc., San Diego, CA, USA) using standard manufacturer's protocols. The 450K array includes more than 450,000 CpGs and is enriched for genic regions, covering 99% of all genes. DNA methylation data pre-processing, including quality control (QC) and normalization, was conducted per cohort using study-specific methods. In all cohorts, DNA methylation levels were quantified as β-values, which range from zero to one, and indicate the proportion of DNA strands in a sample methylated at a specific CpG. Detailed information about cohort specific DNA extraction, bisulfite conversion, DNA methylation profiling, normalization and QC is described in detail in the **Supplementary Text**.

Statistical analysis

Epigenome-wide association analysis

In all cohorts, the association between depressive symptoms and CpG sites was assessed using linear regression analysis in the R software. In the regression analysis, DNA methylation β-value at each CpG site was specified as the dependent variable and the depressive symptoms/depression as the predictor of interest. Association analysis was adjusted for age[42], sex[43], smoking[35] (assessed at the time of blood sampling for methylation), methylation batch effects, white blood cell composition (imputed or directly measured), principal components estimated using genome-wide genotype data to control for population stratification and

familial relationships when required. Cohort specific details of these analyses are provided in the **Supplementary Text**. Further, sensitivity analysis was performed by adjusting the initial model for antidepressant medication use at the time of sample collection.

To pool the results from independent studies we performed sample-size weighted meta-analysis in METAL[44]. We chose the 'sample-size weighted' method because of the differences in the measurement scales of depressive symptoms across studies. A drawback of using sample-size weighted method is that no pooled effect estimates are generated. To obtain pooled effect estimates we additionally performed inverse-variance weighted meta-analysis for the top sites in cohorts that used CES-D 20 item scale for the assessment of depressive symptoms. CpG sites missing in more than three of the participating cohorts were removed. In total, 484,516 probes were tested for association, giving a Bonferroni-corrected genome-wide significance threshold of $0.05/484{,}516 = 1.03\times10^{-7}$. All CpG sites suggestive of association (*p-value* $\leq 10^{-5}$) were tested for association in the independent replication cohorts using the same model as used in the discovery EWAS. Finally, a sample size weighted meta-analysis was performed for all cohorts included in the discovery and replication phases in METAL. To evaluate the contribution of each study to the association results we generated posterior probabilities of the effects in each study (M-values) using the METASOFT package[45]. M-value and Forest plots for z-scores were generated using custom-made scripts in R. For annotating CpG sites we used the annotation provided by Illumina and the UCSC database (GRCh37/hg19).

Gene expression analyses

To evaluate the downstream effects of the three identified CpG sites in blood we used the BIOS database to search for eQTM[46]. To evaluate whether the expression of the genes associated

with or harbored the significant methylation site is associated with major depression (also smoking and inflammation to check specificity) we used the MetaXcan package[47,48]. MetaXcan associates the expression of the genes with the phenotype by integrating functional data generated by large-scale efforts, e.g., Genotype-Tissue Expression (GTEx) with that of the GWAS. MetaXcan is trained on transcriptome models in 44 human tissues from GTEx and is able to estimate their tissue-specific effect on phenotypes from GWAS. We used the GTEx-V6p-HapMap-2016-09-08 database and the publicly available GWAS datasets of major depression[49], and C-reactive protein[50] and smoking[51], which represent important potential confounders in the present study.

Causal inference analysis

To help infer causal relationships, we studied the *cis*-SNPs identified by the BIOS consortium[46] as instrumental variables for the CpG sites as proposed by Smith et al.[52]. We checked the association of these cis-SNPs with depression, smoking and inflammation in the published GWAS of these traits. Similarly, we checked whether the single nucleotide polymorphisms (SNPs) associated with inflammation (CRP levels)[50], smoking[51] and depression[7] were associated with the identified CPG sites using the BIOS consortium database. We chose smoking and inflammation as these are highly correlated with both depression and DNA methylation and thus could influence the relationship between depression and DNA methylation.

Results

The mean age in the discovery cohorts ranged from 52.4 years (SD=8.1) in GS to 79.1 years (SD=0.57) in LBC1921. Forty-eight percent of the total discovery sample were female. The replication cohort comprised 74% women and had an average age of 60.3 years (SD=6.4) (**Table 1**).

Epigenome-wide association analysis

In the meta-analysis of depressive symptoms of European ancestry, we identified one CpG site on chromosome 14q32.32 (cg04987734, *CDC42BPB*, *p-value*=4.93×10^{-8}, *n*=7948) that passed the Bonferroni threshold for significance (**Table 2, eFigure 1**). Further, suggestive association was observed at 19 additional CpG sites (**Table 2**). Adjusting for anti-depressive medication use did not meaningfully change the results (**eTable 1**). No inflation in the test statistic was observed (Lambda=1.03, **eFigure 2**). We tested all 20 CpG sites for association in the replication sample. The top CpG site from the discovery (cg04987734) showed nominal association (*p-value*<0.05, *n*=3308) with depressive symptoms in the validation data set (**Table 2**). In addition, significant association of a CpG site (cg12325605; *p-value*=9.17×10^{-05}, *n*=3308, **Table 2**) annotated to the *ARHGEF3* gene with depressive symptoms was observed in the replication sample.

Meta-analysis of discovery and replication cohorts showed a significant association of both cg04987734 (*p-value*=1.57×$10^{-08,}$ *n*=11256) and cg12325605 (*p-value*=5.24×10^{-09}, *n*=11256) with depressive symptoms (**Table 2; Figures 1 & 2**). Also, an additional intergenic CpG site (cg14023999; *p-value*=5.99×10^{-08}, *n*=11256) at chromosome 15q26.1 locus showed genome-wide significant association with depressive symptoms (**eTable 2, eFigures 3 & 4**). The

independent contributions of each cohort to the association signals of the three CpG are depicted in **eFigure 5** and also provided in **eTable 3**. For all three CpG sites the association signals were not driven by a single cohort but appeared to be linearly related to the sample size, suggesting stronger association in larger studies (**eFigure 5**). Pooled effect estimates in cohorts that used CES-D scale suggest that a 1-unit increase in CES-D score increases methylation by 0.05% at cg04987734, 0.04% at cg12325605, and 0.03% at cg14023999.

Gene expression analyses

Cg04987734 was significantly associated with increased expression of CDC42BPB gene (FDR *p-value*=7.7×10^{-04}, *n*=2101) and cg14023999 was significantly associated with decreased expression of SEMA4B (FDR *p-value*=4.7×10^{-03}, *n*=2101) in blood (**eTable 4**). No significantly associated gene expression probes were identified for cg12325605 in blood. Further, the predicted expression of *CDC42BPB* gene in the brain (basal ganglia) (effect=0.14, *p-value*=2.7×10^{-03}) and of *ARHGEF3* in fibroblasts (effect=-0.48, *p-value*=9.8×10^{-04}) was associated with major depression (**eTable 5**). No association was observed with either smoking or inflammation.

Blood & brain correlation

We checked the correlation between methylation in blood and various brain regions at the three identified sites using a web-based tools, BECon[18] and a blood brain DNA methylation comparison tool (http://epigenetics.essex.ac.uk/bloodbrain/). BECon showed strong correlation between blood and brain DNA methylation, e.g. methylation at cg04987734 in the CDC42BPB gene was highly correlated (r=0.81) between blood and the Brodmann area 7 that spans the medial and lateral walls of the parietal cortex (**eFigure 6**). Methylation at the other two sites

was negatively correlated with methylation in the Brodmann area 10 than spans anterior prefrontal cortex (cg12325605, r=-0.39; cg14023999, r=-0.42) suggesting strong but reverse methylation patterns in blood and brain (**eFigures 7 & 8**). However, the blood brain DNA methylation comparison tool that compares DNA methylation between blood and prefrontal cortex, entorhinal cortex, superior temporal gyrus and cerebellum, showed only modest correlations. For instance, methylation in blood at cg04987734 showed the strongest correlation with methylation in superior temporal gyrus (r = 0.18; http://epigenetics.essex.ac.uk/bloodbrain/?probenameg=cg04987734), while methylation in blood at cg12325605 (http://epigenetics.essex.ac.uk/bloodbrain/?probenameg=cg12325605) and cg14023999 (http://epigenetics.essex.ac.uk/bloodbrain/?probenameg=cg14023999) showed strongest correlation with methylation in cerebellum (r = 0.16 & 0.19 respectively). Nevertheless, the findings from the two databases suggest some degree of correlation between methylation in blood and methylation in brain for the three identified CpG sites.

Causal inference

In the BIOS database we identified two cis-SNPs for cg04987734 and 4 cis-SNPs for cg12325605 (**eTable 6**) and none for cg14023999. We took the most significant cis-SNP as the proxy for the CpG sites if available. For cg04987734 we used rs751837 as a proxy and for cg12325605 we used rs3821412 as a proxy (top cis-SNP rs9880418 was not available in the GWAS of depression, smoking or inflammation). Rs751837 was suggestively associated with major depression (p-value=0.07; albeit in opposite direction) (**eTable 7**). Rs3821412 was not associated with any of the three tested phenotypes. None of the SNPs associated with depression, inflammation or smoking was associated with any of the three CpG sites.

Discussion

In this large-scale EWAS of depressive symptoms, we identified methylation at three CpG sites (cg04987734, cg12325605 and cg14023999) associated with depressive symptoms in the middle-aged and elderly persons. Cg04987734 is annotated to the *CDC42BPB* gene, cg12325605 to the *ARHGEF3* gene, and cg14023999 lies in an intergenic region on chromosome 15q26.1 locus. The predicted expression of *CDC42BPB* and *ARHGEF3* genes associate with major depression in brain and fibroblasts respectively.

CDC42BPB (CDC42 Binding Protein Kinase Beta) encodes a member of the serine/threonine protein kinase family, which is an important downstream effector of CDC42 and plays a role in the regulation of cytoskeleton reorganization, cell migration and regulation of neurite outgrowth[53]. CDC42BPB is highly expressed in the brain https://www.proteinatlas.org/ENSG00000198752-CDC42BPB/tissue. Hyper-methylation of cg04987734 has been associated with increased expression of CDC42BPB in blood[46]. Interestingly, methylation levels at this CpG site (cg04987734) in *CDC42BPB* gene were also previously associated with C-reactive protein (CRP) levels in blood[34]; and smoking[35]. In our study, however, we adjusted for smoking in the regression model; therefore the association between depression and DNA methylation of this CpG site may be independent of smoking habits. Also our causal inference analyses provide no support for the possibility that smoking or inflammation explained the observed association with depressive symptoms nor the predicted expression of the gene showed an association with smoking or inflammation.

ARHGEF3 encodes for Rho Guanine Nucleotide Exchange Factor 3 protein. The gene is highly expressed (https://www.proteinatlas.org/ENSG00000163947-ARHGEF3/tissue) in adrenal glands, brain and uterus. Both *ARHGEF3* and *CDC42BPB* are co-expressed with several members of the Rho subfamily (RHOA, RHOB and RHOC; **eFigures 9 & 10**) of the Rho GTPase family that also includes CDC42[54]. The Rho family of GTPases is a family of small signaling G proteins involved in p75 neurotrophin receptor (p75NTR)-mediated signaling[55] and semaphorin signaling pathways[56]. P75NTR is a transmembrane receptor for neurotrophic factors of the neurotrophin family, which includes the brain-derived neurotrophic factor (BDNF)[57]. P75NTR is widely expressed in the developing central and peripheral nervous system during the period of synaptogenesis and developmental cell death[58]. Both p75NTR and semaphorins are implicated in axon guidance[59,60]. In this context, the third associated CpG site cg14023999 that lies in an intergenic region on chromosome 15q26.1 is also interesting. Cg14023999 is associated with decreased expression of *SEMA4B* gene in blood. *SEMA4B* encodes for Semaphorin 4B protein. Sema4B is believed to function through a direct interaction with post-synaptic density protein PSD-95[61] to promote synapse maturation[61-63]. The knock-down of Sema4B causes a decrease in GABAergic synapse number[62] suggesting a role in the assembly of excitatory and inhibitory postsynaptic specializations[63]. Previously cg14023999 was found to be significantly correlated with Parkinson's disease[64] and significant association of a CpG site in *SEMA4B* was observed in individuals with schizophrenia carrying the 22q11.2 deletion[65]. These findings point towards a functional of *SEMA4B* in neuro-psychiatric disorders. When comparing our findings with that of the previous EWAS of depression, we did not find an overlap. These studies were small (<100 individuals) and did not report reproducible results[20].

To summarize, we report the first EWAS of depressive symptoms. We identified and replicated association of two methylation sites in the genome with depressive symptoms. A third site was identified in the meta-analysis of discovery and replication cohorts, which requires further replication. All three findings point towards axon guidance as the common disrupted pathway in depression (http://www.genome.jp/kegg/pathway/hsa/hsa04360.html). Our findings provide new insights into the molecular mechanisms underlying the complex pathophysiology of depression.

Strengths and limitations of the study

This is the largest epigenome-wide study of depressive symptoms reported to date. Our major strength is the sample size that enabled detection of a replicable epigenome-wide significant locus, which suggests that in blood, DNA methylation signatures associated with depression may be subtle and will require large samples to be detected. Using peripheral blood tissue for DNA methylation profiling is a limitation of this study, as DNA methylation is known to be tissue specific[66]. While peripheral blood is not considered to be the most relevant tissue for the pathophysiology of depression, some sites show correlated methylation profiles between-tissues[15,66]. The three sites identified in our study show some degree of correlation between methylation in blood and various brain regions. Second, while replication in African-American samples suggests that some depressive symptoms related differences in DNA methylation may be similar across ethnicities[67-69], it may also have resulted in false negatives due to different genetic background. Third, in these analyses we mostly used quantitative measures of depressive symptoms. Quantitative endo-phenotypes provide powerful alternatives for several complex outcomes, for example, hypertension[70]. This is likely to be especially true for a trait

such as depressive symptoms, for which the severity and duration of illness can be highly heterogeneous[22]. Genome-wide studies of depressive traits, using quantitative endo-phenotypes, have been suggested to improve statistical power[22]. However, the use of different phenotypic measures by different cohorts means that there may be some loss of statistical power due to the heterogeneity in the phenotype assessment. Nevertheless, the top three sites in our study were robustly associated with depressive symptoms independent of the depressive symptom measure used. Fourth, although we adjusted for potential confounders, the possibility of residual confounding cannot be excluded. Antidepressant medication indicates treated depression but itself may result in epigenetic modifications involved in depression pathophysiology[71]. Antidepressants can thus mediate or confound the relation between DNA methylation and depression. However, in sensitivity analysis additionally adjusted for antidepressant medication, our results did not change. Fifth, most cohorts included in this EWAS are cohorts of elderly persons. The etiology of depression is more heterogeneous in elderly people than in younger and often hidden behind somatic symptoms, either because of somatization of the disorder or because of accentuation of symptoms of concomitant illness[72]. This may affect the generalizability of the results to younger populations. Finally, we made an attempt to disentangle cause and consequence using SNPs associated with the identified CpG sites and depression, inflammation and smoking as instrumental variables. The results did not support a causative role yet the association of the predicted gene expression of CDC42BPB in brain and ARHGEF3 in fibroblasts with major depression does suggest a possible causal role of the regulatory effects of these genes[47].

Acknowledgments

The authors acknowledge the essential role of the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) Consortium in development and support of this manuscript. CHARGE members include the Netherland's Rotterdam Study (RS), Framingham Heart Study (FHS), Cardiovascular Health Study (CHS), the Helsinki Birth Cohort Study (HBCS), Cooperative Health Research in the Augsburg Region (KORA), Lothian Birth Cohort (LBC), Generation Scotland Study (GS), Atherosclerosis Risk in Communities Study (ARIC), and the Women's Health Initiative - Epigenetic Mechanisms of PM-Mediated Cardiovascular disease (WHI-EMPC). Detailed acknowledgments are included in the supplementary material.

Henning Tiemeier and Najaf Amin have full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

COMPETING FINANTIAL INTERESTS

The authors declare no competing financial interests.

References

1. Lepine JP, Briley M. The increasing burden of depression. *Neuropsychiatr Dis Treat.* 2011;7(Suppl 1):3-7.
2. Rodda J, Walker Z, Carter J. Depression in older adults. *BMJ.* 2011;343:d5219.
3. Conwell Y, Duberstein PR, Cox C, Herrmann JH, Forbes NT, Caine ED. Relationships of age and axis I diagnoses in victims of completed suicide: a psychological autopsy study. *Am J Psychiatry.* 1996;153(8):1001-1008.
4. Hidaka BH. Depression as a disease of modernity: explanations for increasing prevalence. *J Affect Disord.* 2012;140(3):205-214.
5. Lohoff FW. Overview of the genetics of major depressive disorder. *Curr Psychiatry Rep.* 2010;12(6):539-546.
6. Gatz M, Pedersen NL, Plomin R, Nesselroade JR, McClearn GE. Importance of shared genes and shared environments for symptoms of depression in older adults. *J Abnorm Psychol.* 1992;101(4):701-708.
7. Wray NR, Sullivan PF. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv.* 2017.
8. CONVERGE. consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature.* 2015;523:588–591.
9. Hek K, Demirkan A, Lahti J, et al. A genome-wide association study of depressive symptoms. *Biol Psychiatry.* 2013;73(7):667-678.
10. Direk N, Williams S, Smith JA, et al. An Analysis of Two Genome-wide Association Meta-analyses Identifies a New Locus for Broad Depression Phenotype. *Biol Psychiatry.* 2017;82(5):322-329.
11. Agid O, Kohn Y, Lerer B. Environmental stress and psychiatric illness. *Biomed Pharmacother.* 2000;54(3):135-141.
12. Berk M, Williams LJ, Jacka FN, et al. So depression is an inflammatory disease, but where does the inflammation come from? *BMC Med.* 2013;11:200.
13. Dahl C, Guldberg P. DNA methylation analysis techniques. *Biogerontology.* 2003;4(4):233-250.
14. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. *Trends Genet.* 2002;18(7):348-351.
15. Lokk K, Modhukur V, Rajashekar B, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol.* 2014;15(4):r54.
16. Byrne EM, Carrillo-Roa T, Henders AK, et al. Monozygotic twins affected with major depressive disorder have greater variance in methylation than their unaffected co-twin. *Transl Psychiatry.* 2013;3:e269.
17. Latham KE. Stage-specific and cell type-specific aspects of genomic imprinting effects in mammals. *Differentiation.* 1995;59(5):269-282.
18. Edgar RD, Jones MJ, Meaney MJ, Turecki G, Kobor MS. BECon: A tool for interpreting DNA methylation findings from blood in the context of brain. *bioRxiv.* 2017.
19. Bakusic J, Schaufeli W, Claes S, Godderis L. Stress, burnout and depression: A systematic review on DNA methylation mechanisms. *J Psychosom Res.* 2017;92:34-44.
20. Januar V, Saffery R, Ryan J. Epigenetics and depressive disorders: a review of current progress and future directions. *Int J Epidemiol.* 2015;44(4):1364-1387.
21. Nagy C, Suderman M, Yang J, et al. Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Mol Psychiatry.* 2015;20(3):320-328.
22. van der Sluis S, Posthuma D, Nivard MG, Verhage M, Dolan CV. Power in GWAS: lifting the curse of the clinical cut-off. *Mol Psychiatry.* 2013;18(1):2-3.

23. Ayuso-Mateos JL, Nuevo R, Verdes E, Naidoo N, Chatterji S. From depressive symptoms to depressive disorders: the relevance of thresholds. *Br J Psychiatry.* 2010;196(5):365-371.
24. Meeks TW, Vahia IV, Lavretsky H, Kulkarni G, Jeste DV. A tune in "a minor" can "b major": a review of epidemiology, illness course, and public health implications of subthreshold depression in older adults. *J Affect Disord.* 2011;129(1-3):126-142.
25. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991;1(3):263-276.
26. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol.* 1979;110(3):281-290.
27. Eriksson JG, Forsen T, Tuomilehto J, Osmond C, Barker DJ. Early growth and coronary heart disease in later life: longitudinal study. *BMJ.* 2001;322(7292):949-953.
28. Wichmann HE, Gieger C, Illig T, Group MKS. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen.* 2005;67 Suppl 1:S26-30.
29. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol.* 2012;41(6):1576-1584.
30. Deary IJ, Gow AJ, Taylor MD, et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr.* 2007;7:28.
31. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol.* 2017;32(9):807-850.
32. Smith BH, Campbell A, Linksted P, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol.* 2013;42(3):689-700.
33. Marioni RE, McRae AF, Bressler J, et al. Meta-analysis of epigenome-wide association studies of cognitive abilities. *Mol Psychiatry.* 2018.
34. Ligthart S, Marzi C, Aslibekyan S, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* 2016;17(1):255.
35. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet.* 2016;9(5):436-447.
36. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* 1989;129(4):687-702.
37. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials.* 1998;19(1):61-109.
38. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging.* 1997;12(2):277-287.
39. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67(6):361-370.
40. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606-613.
41. First MB, Spitzer, Robert L, Gibbon Miriam, and Williams, Janet B.W. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P) November 2002.
42. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14(10):R115.
43. Singmann P, Shem-Tov D, Wahl S, et al. Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics Chromatin.* 2015;8:43.

44. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
45. Han B, Eskin E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* 2012;8(3):e1002555.
46. Bonder MJ, Luijk R, Zhernakova DV, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 2017;49(1):131-138.
47. Barbeira A, Dickinson SP, Torres JM, et al. Integrating tissue specific mechanisms into GWAS summary results. *bioRxiv.* 2017.
48. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet.* 2015;47(9):1091-1098.
49. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry.* 2013;18(4):497-511.
50. Dehghan A, Dupuis J, Barbalic M, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation.* 2011;123(7):731-738.
51. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42(5):441-447.
52. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol.* 2012;41(1):161-176.
53. Chen XQ, Tan I, Leung T, Lim L. The myotonic dystrophy kinase-related Cdc42-binding kinase is involved in the regulation of neurite outgrowth in PC12 cells. *J Biol Chem.* 1999;274(28):19901-19905.
54. Etienne-Manneville S, Hall A. Rho GTPases in cell biology. *Nature.* 2002;420(6916):629-635.
55. DeGeer J, Lamarche-Vane N. Rho GTPases in neurodegeneration diseases. *Exp Cell Res.* 2013;319(15):2384-2394.
56. Liu BP, Strittmatter SM. Semaphorin-mediated axonal guidance via Rho-related G proteins. *Curr Opin Cell Biol.* 2001;13(5):619-626.
57. Woo NH, Teng HK, Siao CJ, et al. Activation of p75NTR by proBDNF facilitates hippocampal long-term depression. *Nat Neurosci.* 2005;8(8):1069-1077.
58. Davies AM. Nerve growth factor synthesis and nerve growth factor receptor expression in neural development. *Int Rev Cytol.* 1991;128:109-138.
59. Fiore R, Puschel AW. The function of semaphorins during nervous system development. *Front Biosci.* 2003;8:s484-499.
60. Schecterson L, Bothwell M. An all-purpose tool for axon guidance. *Neuroscience.* 2008;1(47):Pe50.
61. Burkhardt C, Muller M, Badde A, Garner CC, Gundelfinger ED, Puschel AW. Semaphorin 4B interacts with the post-synaptic density protein PSD-95/SAP90 and is recruited to synapses through a C-terminal PDZ-binding motif. *FEBS Lett.* 2005;579(17):3821-3828.
62. Paradis S, Harrar DB, Lin Y, et al. An RNAi-based approach identifies molecules required for glutamatergic and GABAergic synapse development. *Neuron.* 2007;53(2):217-232.
63. Pasterkamp RJ, Giger RJ. Semaphorin function in neural plasticity and disease. *Curr Opin Neurobiol.* 2009;19(3):263-274.
64. Chuang YH, Paul KC, Bronstein JM, Bordelon Y, Horvath S, Ritz B. Parkinson's disease is associated with DNA methylation levels in human blood and saliva. *Genome Med.* 2017;9(1):76.

65. Starnawska A, Hansen CS, Sparso T, et al. Differential DNA methylation at birth associated with mental disorder in individuals with 22q11.2 deletion syndrome. *Transl Psychiatry.* 2017;7(8):e1221.
66. Bagot RC, Labonte B, Pena CJ, Nestler EJ. Epigenetic signaling in psychiatric disorders: stress and depression. *Dialogues Clin Neurosci.* 2014;16(3):281-295.
67. Klengel T, Pape J, Binder EB, Mehta D. The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology.* 2014;80:115-132.
68. Galanter JM, Gignoux CR, Oh SS, et al. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *Elife.* 2017;6.
69. Kader F, Ghai M. DNA methylation-based variation between human populations. *Mol Genet Genomics.* 2017;292(1):5-35.
70. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011;478(7367):103-109.
71. Lisoway AJ, Zai CC, Tiwari AK, Kennedy JL. DNA Methylation and Clinical Response to Antidepressant Medication in Major Depressive Disorder: A Review and Recommendations. *Neurosci Lett.* 2017.
72. Gottfries CG. Is there a difference between elderly and younger patients with regard to the symptomatology and aetiology of depression? *Int Clin Psychopharm.* 1998;13:S13-S18.

Table 1. Descriptive statistics of the discovery and replication cohorts.

Study	Ethnicity	N	Female (%)	Mean Age (SD)	Current smokers (%)	Depressive symptoms	Antidepressant medication use (%)
Discovery (N = 7948)							
CHS	European	323	194 (60.1)	75.6 (5.2)	173 (53.6)	CESD[1] (10 item)	19 (5.9)
FHS	European	2722	1508 (53.6)	58.5 (11.6)	948 (34.8)	CESD[2] (20 item)	251 (16.1)
HBCS	European	122	0 (0)	65.2 (2.7)	24 (19.7)	CESD[2] (20 item)	11 (9.0)
KORA	European	1727	882 (51.1)	61.0 (8.9)	250 (14.5)	PHQ-9[3]	82 (4.7)
LBC 1921	European	432	261 (60.4)	79.1 (0.6)	194 (44.9)	HADS[4]	15 (3.5)
LBC 1936	European	916	452 (49.3)	69.6 (0.8)	504 (55)	HADS[4]	30 (3.3)
RS III	European	722	391 (54.2)	59.8 (8.1)	167 (23.1)	CESD[2] (20 item)	38 (5.3)
RS BIOS	European	757	319 (42.1)	67.6 (5.9)	78 (10.3)	CESD[2] (20 item)	51 (6.7)
GS[a]	European	227	151 (64.5)	52.4 (8.1)	46 (19.7)	SCID[5]	44 (18.8)
Total		7948	4158 (48.4)	65.4 (5.8)	2384 (30.6)	-	541 (8.1)
Replication (N = 3308)							
ARIC	African	2297	1445 (63)	56.1 (5.7)	584 (25.4)	21-MQ[6]	74 (3.3)
WHI-EMPC	European	1011	1011 (100)	64.6 (7.1)	509 (50.3)	CES-D/DIS[7]	61 (6.0)
Total		3308	2456 (74.2)	60.3 (6.4)	1093 (37.9)	-	135 (4.7)

Characteristics are depicted as mean (SD), unless otherwise specified.

CHC Cardiovascular health cohort, ***FHS*** Framingham Heart Study, ***HBCS*** Helsinki Birth Cohort Study, ***KORA*** Cooperative Health Research in the Augsburg Region, ***LBC*** Lothian Birth Cohort, ***RS*** Rotterdam Study, ***GS*** Generation Scotland Study, and [a] CASE-CONTROL STUDY, ***ARIC*** Atherosclerosis Risk in Communities Study and ***WHI-EMPC*** the Women's Health Initiative - Epigenetic Mechanisms of PM-Mediated Cardiovascular disease***;*** "(in brackets we state number of item of the questionnaires)".

1 Irwin, M. et al. 1999. [2] Radloff, LS. et al. 1977. [3] Kroenke, K. et al. 2001. [4] Zigmond, AS. et al. 1983. [5] First, MB. et al. 1996. [6] Wattanakit,K. et al. 2005. [7] Burnam, MA. et al. 1988.

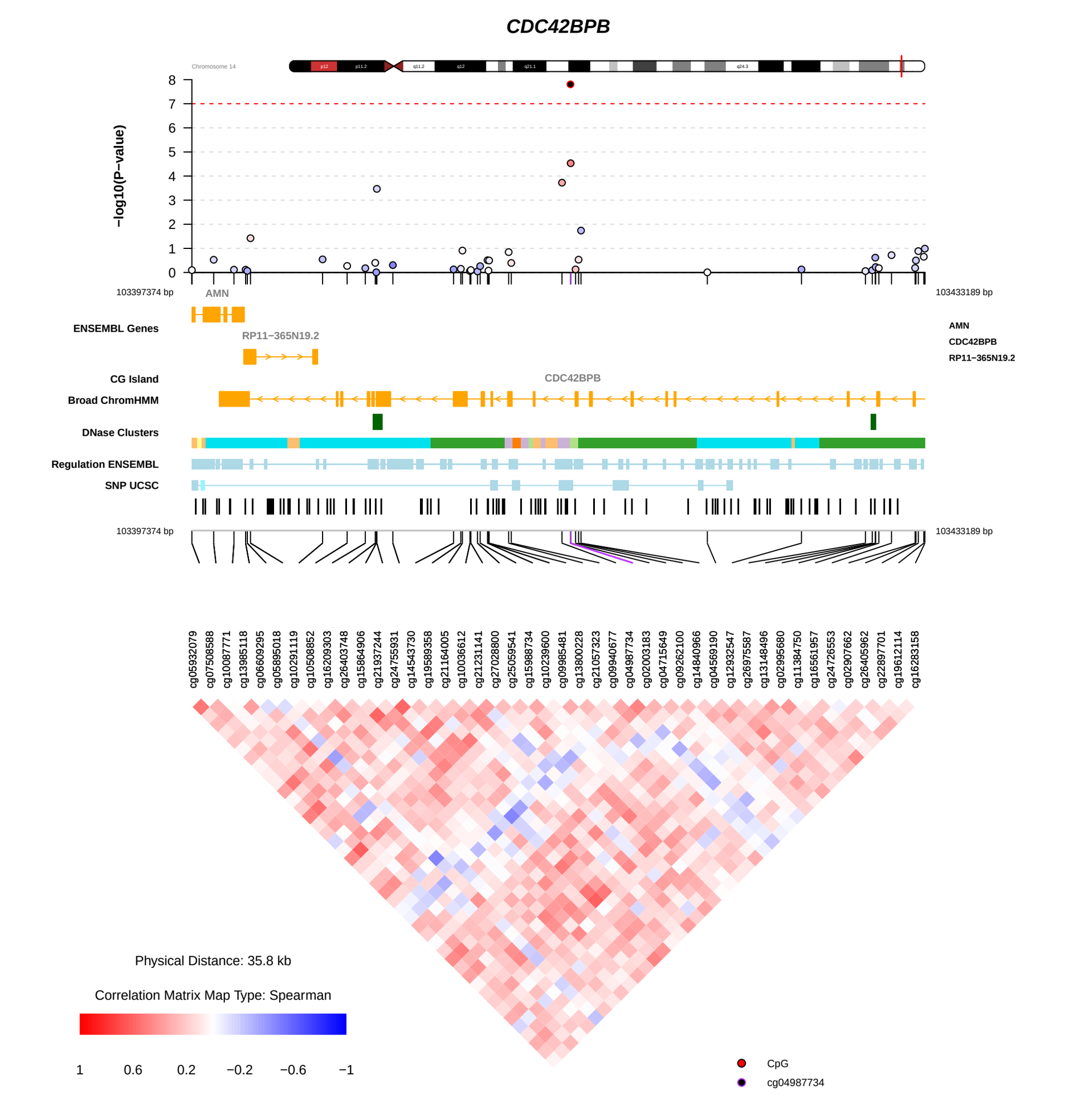
Table 2. Top DNA methylation sites associated with depressive symptoms in the discovery EWAS.

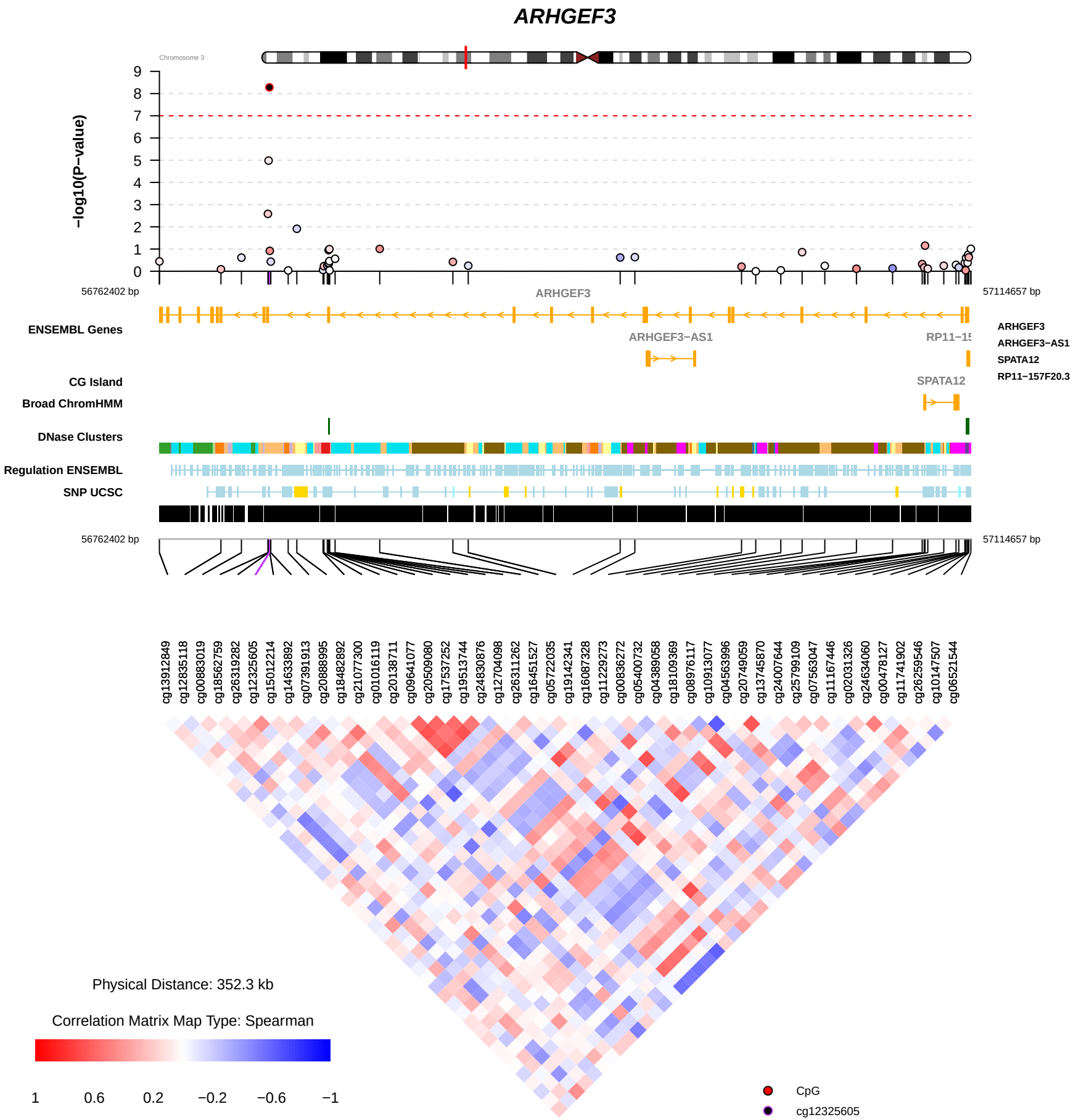
				Discovery (N=7948)	Replication (N=3308)	Meta-analysis (N=11256)
CpG site ID	**Chr**	**Location**	**Gene symbol**	**P-value**	**P-value**	**P-value**
cg04987734	14	103415873	*CDC42BPB*	**4.93×10^{-8}**	4.82×10^{-02}	**1.57×10^{-08}**
cg07012687	17	80195180	*SLC16A3*	3.47×10^{-7}	1.58×10^{-01}	4.45×10^{-06}
cg08796240	16	70733832	*VAC14*	7.43×10^{-7}	2.56×10^{-01}	1.80×10^{-06}
cg06096336	2	231989800	*PSMD1; HTR2B*	8.06×10^{-7}	3.01×10^{-01}	2.51×10^{-06}
cg16745930	10	100220809	*HPSE2*	1.34×10^{-6}	4.01×10^{-01}	6.26×10^{-06}
cg09849319	5	1494983	*LPCATI1*	1.81×10^{-6}	4.64×10^{-01}	1.04×10^{-04}
cg17237086	22	40814966	*MKL1*	3.44×10^{-6}	2.51×10^{-01}	6.10×10^{-06}
cg03985718	2	105924245	*TGFBRAP1*	3.61×10^{-6}	8.54×10^{-01}	6.53×10^{-05}
cg21098005	20	44538605	*PLTP*	4.36×10^{-6}	9.60×10^{-01}	1.01×10^{-04}
cg16466652	19	6271960	*MLLT1*	4.39×10^{-6}	3.97×10^{-01}	1.57×10^{-05}
cg07884764	11	64107517	*CCDC88B*	5.03×10^{-6}	9.99×10^{-01}	1.25×10^{-04}
cg01541347	7	4729920	*FOXK1*	5.64×10^{-6}	3.77×10^{-01}	8.46×10^{-04}
cg02341197	21	34185927	*C21or62*	5.84×10^{-6}	2.02×10^{-01}	6.80×10^{-06}
cg01947751	3	196728969	-	6.23×10^{-6}	6.63×10^{-01}	3.68×10^{-04}
cg13747876	17	80195402	*SLC16A3*	6.32×10^{-6}	1.04×10^{-01}	2.93×10^{-06}
cg12764201	1	105101123	*CORT; APITD1*	7.15×10^{-6}	7.20×10^{-01}	7.29×10^{-05}
cg08295111	5	133866097	*PHF15*	7.87×10^{-6}	5.76×10^{-01}	5.64×10^{-04}
cg18030453	3	45506216	*LARS2*	9.16×10^{-6}	3.87×10^{-03}	1.20×10^{-07}
cg12325605	3	56810151	*ARHGEF3*	9.62×10^{-6}	**9.17×10^{-05}**	**5.24×10^{-09}**
cg23282441	10	73533927	*C10orf54; CDH23*	9.69×10^{-6}	1.77×10^{-01}	8.63×10^{-06}

Figure legends

Figure 1: Regional association plot for the top CpG site cg04987734. The horizontal axis depicts the position in base pair (hg19) for the entire *CDC42BPB* gene region. The Vertical axis indicates the strength of association in terms of negative logarithm of the association *p*-value. Each circle represents CpG site. Red dashed line indicates the genome-wide significance threshold. Below the horizontal axis the figure shows the regulatory information and correlation matrix of other CpG sites in the region with the top hit. Color intensity marks the strength of the correlation and color the direction of the correlation. Figure is made using web-based plotting tool and R-based package "CoMET" (http://comet.epigen.kcl.ac.uk:3838/coMET1/).

Figure 2: Regional association plot for the top CpG site cg12325605. The horizontal axis depicts the position in base pair (hg19) for the entire *ARHGEF3* gene region. The Vertical axis indicates the strength of association in terms of negative logarithm of the association *p*-value. Each circle represents CpG site. Red dashed line indicates the genome-wide significance threshold. Below the horizontal axis the figure shows the regulatory information and correlation matrix of other CpG sites in the region with the top hit. Color intensity marks the strength of the correlation and color the direction of the correlation. Figure is made using web-based plotting tool and R-based package "CoMET" (http://comet.epigen.kcl.ac.uk:3838/coMET1/).


CDC42BPB
Chromosome 14
p12
p11.2
q11.2
q12
q21.1
q24.3
−log10(P−value)
0
1
2
3
4
5
6
7
8
103397374 bp
103433189 bp
AMN
ENSEMBL Genes
RP11−365N19.2
AMN
CDC42BPB
RP11−365N19.2
CG Island
CDC42BPB
Broad ChromHMM
DNase Clusters
Regulation ENSEMBL
SNP UCSC
103397374 bp
103433189 bp
cg05932079
cg07508588
cg10087771
cg13985118
cg06609295
cg05895018
cg10291119
cg10508852
cg16209303
cg26403748
cg15864906
cg21937244
cg24755931
cg14543730
cg19589358
cg21164005
cg10036612
cg21231141
cg27028800
cg25059541
cg15988734
cg10239600
cg09985481
cg13800228
cg21057323
cg09940677
cg04987734
cg02003183
cg04715649
cg09262100
cg14840966
cg04569190
cg12932547
cg26975587
cg13148496
cg02995680
cg11384750
cg16561957
cg24726553
cg02907662
cg26405962
cg22897701
cg19612114
cg16283158
Physical Distance: 35.8 kb
Correlation Matrix Map Type: Spearman
1
0.6
0.2
−0.2
−0.6
−1
CpG
cg04987734


ARHGEF3
Chromosome 3
−log10(P−value)
9
8
7
6
5
4
3
2
1
0
56762402 bp
57114657 bp
ARHGEF3
ENSEMBL Genes
ARHGEF3−AS1
RP11−15
ARHGEF3
ARHGEF3−AS1
SPATA12
RP11−157F20.3
CG Island
SPATA12
Broad ChromHMM
DNase Clusters
Regulation ENSEMBL
SNP UCSC
56762402 bp
57114657 bp
cg13912849
cg12835118
cg00883019
cg18562759
cg26319282
cg12325605
cg15012214
cg14633892
cg07391913
cg20888995
cg18482892
cg21077300
cg01016119
cg20138711
cg09641077
cg20509080
cg17537252
cg19513744
cg24830876
cg12704098
cg26311262
cg16451527
cg05722035
cg19142341
cg16087328
cg11229273
cg00836272
cg05400732
cg04389058
cg18109369
cg08976117
cg10913077
cg04563996
cg20749059
cg13745870
cg24007644
cg25799109
cg07563047
cg11167446
cg02031326
cg24634060
cg00478127
cg11741902
cg26259546
cg10147507
cg06521544
Physical Distance: 352.3 kb
Correlation Matrix Map Type: Spearman
1
0.6
0.2
−0.2
−0.6
−1
CpG
cg12325605

Supplementary Information

"DNA methylation signatures of depressive symptoms in middle-aged and elderly persons identified in a large multi-ethnic meta-analysis of epigenome-wide studies"

Olivera Jovanova[1†], PhD, Ivana Nedeljkovic[1†], MD, Spieler Derek[2, 3†], MD, Rosie M. Walker[4, 5†], PhD, Chunyu Liu[6, 7, 8†], PhD, Michelle Luciano[4, 9†], PhD, Jan Bressler[10], PhD, Jennifer Brody[11], BA, Amanda J. Drake[12], PhD, Kathryn L. Evans[4, 5], PhD, Rahul Gondalia[13], MPH, Sonja Kunze[2, 14], PhD, Brigitte Kuhnel[2, 14], MSc, Jari Lahti[15], PhD, Rozenn N. Lemaitre[11], PhD, Riccardo E. Marioni[4, 9], PhD, Brenton Swenson[11, 16], MSc, Jayandra Jung Himali[17], PhD, Hongsheng Wu[18], PhD, Yun Li[19, 20, 21], PhD, Allan F. McRae[22], PhD, Tom C. Russ[4, 23, 24], PhD, James Stewart[13, 25], MSc, Zhiying Wang[10], MSc, Guosheng Zhang[19, 20, 21], MSc, Karl-Heinz Ladwig[2, 3], PhD, Andre G. Uitterlinden[1, 26], PhD, Xiuqing Guo[27], PhD, Annette Peters[2, 14], PhD, Katri Räikkönen[15], PhD, John M. Starr[4, 23], PhD, Melanie Waldenberger[2, 14], PhD, Naomi R. Wray[22], PhD, Whitsel A. Eric[13, 28], MD, Nona Sotoodehnia[11], MD, Sudha Seshadri[6, 29], MD, David J. Porteous[4, 5], PhD, Joyce van Meurs[26], PhD, Thomas H. Mosley[30], PhD, Andrew M. McIntosh[4, 31], PhD, Michael M. Mendelson[6, 7, 32], MD, Daniel Levy[6, 7], MD, Lifang Hou[33], PhD, Johan G. Eriksson[34], PhD, Myriam Fornage[10, 35], PhD, Ian J. Deary[4, 9], PhD, Andrea Baccarelli[36], PhD, Henning Tiemeier[1, 37, 38*], PhD, Najaf Amin[1], PhD.

[1] Department of Epidemiology, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[2] Institute of Epidemiology II, Helmholtz Zentrum München, 85764 Neuherberg, Germany.
[3] Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie des Klinikums Rechts der Isar der TUM, 81675 Munich, Germany.
[4] Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK.
[5] Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, The University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK.
[6] The Framingham Heart Study, Framingham, MA, USA.
[7] The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA.
[8] Boston University School of Public Health, 715 Albany St, Boston, MA, USA.
[9] Psychology, University of Edinburgh, Edinburgh, UK.

[10] Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.
[11] Cardiovascular Health Research Unit, Department of Medicine, University of Washington.
[12] University/British Heart Foundation Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh.
[13] Department of Epidemiology, University of North Carolina. Chapel Hill, NC 27514.
[14] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany.
[15] Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland.
[16] Institute for Public Health Genetics, School of Public Health, University of Washington.
[17] Boston University School of Medicine and of Public Health, The Framingham Study, 72 East Concord Street, B 604, Boston, MA 021118, USA.
[18] Computer Science and Networking, Wentworth Institute of Technology, Boston, MA.
[19] Department of Genetics, University of North Carolina. Chapel Hill, NC 27514.
[20] Department of Biostatistics, University of North Carolina. Chapel Hill, NC 27514.
[21] Department of Computer Science, University of North Carolina. Chapel Hill, NC 27514.
[22] Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia; Queensland Brain Institute, The University of Queensland, Brisbane, Australia.
[23] Alzheimer Scotland Dementia Research Centre, 7 George Square, Edinburgh, UK. EH8 9JZ.
[24] Centre for Dementia Prevention, University of Edinburgh, UK.
[25] Carolina Population Center, University of North Carolina. Chapel Hill, NC 27514.
[26] Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3000 CA, The Netherlands.
[27] The Institute for Translational Genomics and Population Sciences, Department of Pediatrics at Harbor-UCLA Medical Center.
[28] Department of Medicine, University of North Carolina. Chapel Hill, NC 27514.
[29] Boston University School of Medicine,72 East Concord St, B602, Boston, MA 02118.
[30] MIND Center, University of Mississippi Medical Center, Jackson, MS.
[31] Division of Psychiatry, The University of Edinburgh, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, UK.
[32] Department of Cardiology, Boston Children's Hospital, Harvard Medical School, Boston, MA, US.
[33] Northwestern University. 680 N Lake Shore Drive, Suite 1400, Chicago IL 60611.
[34] Department of General Practice and Primary health Care, University of Helsinki, Tukholmankatu 8 B, Helsinki, 00014, Finland.
[35] Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX.
[36] Harvard University. 665 Huntington Ave, Boston, MA 02115.
[37] Department of Child and Adolescent Psychiatry, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[38] Department of Social and Behavioral Science, Harvard TH Chan School of Public Health, Boston, USA.
[†] Equal contributors.
[*] Corresponding author.

***Corresponding author:** Henning Tiemeier, MD, PhD,

Epidemiology, Erasmus MC-University Medical Center Rotterdam, P.O. Box 2040, 3000 CA

Rotterdam, the Netherlands, T:+31 107043489, F:+31 107044657.

Email: h.tiemeier@erasmusmc.nl

Study-specific details

Cardiovascular Health Study Cohort

Description. The CHS is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥65 years conducted across four field centers.(1) The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists. CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease.

DNA methylation sample, measurement, normalization and quality control. DNA methylation was measured on a randomly selected subset of 323 European descent participants from study year. The samples were randomly selected among participants without presence of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack at study baseline or lack of available DNA at study year 5.
Methylation measurements were performed at the Institute for Translational Genomics and Population Sciences at the Harbor-UCLA Medical Center Institute for Translational Genomics and Population Sciences using the Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA). Quality control was performed in in the minfi R package.(2-4) (version 1.12.0, http://www.bioconductor.org/packages/release/bioc/html/minfi.html). Samples with low median intensities of below 10.5 (log2) across the methylated and unmethylated channels, samples with a proportion of probes falling detection of greater than 0.5%, samples with QC probes falling greater than 3 standard deviation from the mean, sex check mismatches, or failed concordance with prior genotyping were removed. Methylation values were normalized using the SWAN quantile normalization method.(3) Since white blood cell proportions were not directly measured in CHS they were estimated from the methylation data using the Houseman method.(5) All association analyses were performed in R using linear mixed models with DNA methylation beta values as the outcome. Analyses adjusted for age, gender, current smoking status and antidepressant medication. Additionally chip, chip row and column were adjusted for as random effects.
CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease. The DNA methylation dataset from CHS can be requested at https://chs-nhlbi.org/node/6222.

Depressive symptoms assessment. Depressive symptoms were assessed during an in-person interview using the Center for Epidemiology Studies-Depression (CES-D) 10 item scale.

Covariates. Smoking and medication were assessed during an in-person interview, at the time of blood assessment for methylation. Smoking was assessed with a standard questionnaire, medications were assessed by an inventory of all prescription medication containers used in the 2 weeks prior to the interview.

Acknowledgments. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. The CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL092111, R01HL105756, R01HL103612, R01HL111089, R01HL116747 and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA) as well as Laughlin Family, Alpha Phi Foundation, and Locke Charitable Foundation. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes

Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Framingham Heart Study (FHS)

Description. The Framingham Heart Study (FHS) is a population-based, prospective study. In 1948, the Original cohort of 5209 individuals were recruited from Framingham, MA.(6) The offspring cohort was recruited in 1971, including 5,124 offspring and spouses of offspring of the FHS Original cohort.(7) Offspring participants underwent examinations every four years (except eight years between the first and the second examinations).(7)

DNA methylation sample, measurement, normalization and quality control. Peripheral whole blood samples were collected from these participants at the eighth examination (2005-2008). Buffy coat fractions were obtained and genomic DNA was extracted using the Gentra Puregene DNA extraction kit (Qiagen, Venlo, Netherlands). Bisulfite conversion of genomic DNA was performed with the EZ DNA Methylation Kit (Zymo Research, Irvine, CA). DNA methylation was measured by *Infinium HumanMethylation450 BeadChip* (Illumina, CA) in two separate labs. All participants provided written consent for genetic study.
QC procedures were performed. Bad samples were excluded if the samples were outliers in multidimensional scaling (MDS) analysis, high missing rate (>1%), poor matching to SNP genotype. Low quality probes were excluded if these probes had high missing rate (>20%), were mapped to multiple locations, had SNP (MAF>5% in EUR 1000G) at CpG site or ≤10 bp of Single Base Extension.(8) About 440,000 probes remain in ~2,600 FHS participants. A total of 1,680 Offspring participants had both DNA methylation and CESD phenotype. The sample characteristics are displayed in Table 1 in the manuscript. The DNA methylation dataset from FHS is available on request to Chunyu Liu e-mail: chunyu.liu@nih.gov.

Depressive symptoms assessment. Depressive symptoms were evaluated using the Center for Epidemiologic Studies Depression Scale (CES-D)(9), a 20-item scale consisting of 4 factors: depressive affect, somatic complaints, positive affect, and interpersonal relations (1). This 20-item, self-report scale, designed for use in community studies, has been validated in studies similar to the Framingham Heart Study . The CES-D scale was self-administered in the offspring cohort. Scores on the CES-D range from 0 to 60, with higher scores reflecting greater depressive symptoms.

Covariates. At each examination, participants underwent a medical history, physical examination, and fasting blood sample collection. Clinical covariates were assessed at the same examination cycle as the blood sample obtained for DNA methylation. Details of the examination assessments have been previously described(10) and are available in detail at http://www.framinghamheartstudy.org/fhs-for-researchers/. Self-reported cigarette smoking status was divided into three categories. Current smokers were defined as those who have smoked at least one cigarette a day within 12 months prior to the blood draw, former smokers were defined as those who had ever smoked at least one cigarette a day, but had stopped at least 12 months prior to the blood draw, and never smokers reported never having smoked. Medication use was ascertained by physicians by detailed review of participant medication lists, and participants were asked to bring medication bottles to the clinic examination for review.

Analysis. We followed the analysis plan. Two models were performed. In order to reduce the batch effects, we analyzed the DNA methylation in each lab: residuals were obtained by regressing a CpG against the batch effects (plate, row and col numbers) and lab-specific principle components that were related to the outcome. Residuals from two labs were put together and then regressed against the CESD, adjusting for covariates and family structure. Plate ID and family structures were treated as random effects in linear mixed models.

Model 1. Resid (CpG ~ Batch effects + PCs) (lab-specific)
Pooled residuals from two labs ~ CESD + Gender + Age + smoking + WBC + family structure

Model 2. Resid (CpG ~ Batch effects + PCs) (lab-specific)
Pooled residuals from two labs ~ CESD + Gender + Age + smoking + WBC + deprx8 + family structure

Acknowledgements. The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD. The analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the Center for Information Technology, National Institutes of Health, Bethesda, MD. This project is also supported by grant contract R01 AG054076**.** Dr. Michael Mendelson was supported by the NIH NHLBI K99-HL136875.

Helsinki Birth Cohort Study (HBCS)

Description. The HBCS comprises 13345 individuals (6370 women and 6975 men), born as singletons between 1934 and 1944 in one of the two main maternity hospitals in Helsinki and who were living in Finland in 1971 when a unique personal identification number was allocated to each member of the Finnish population. The HBCS, which has been described in detail elsewhere, has been approved by the Ethics Committee of the National Public Health Institute. Register data were linked with permission from the Finnish Ministry of Social Affairs and Health and the Finnish National Archives.
In 2001 – 2004 at an average age of 61.5 years (SD = 2.9 and range = 56.7 – 69.8 years), a randomly selected subsample of the cohort comprising 2003 individuals (1075 women and 928 men) was invited to a clinical examination including collection of a blood sample for genetic, epigenetic and biochemical studies and a psychological survey including a measure of depressive symptoms. For 283 participants, extraction of DNA was not successful, or DNA showed gender discrepancy or close relatedness. From the remaining sample of 1720 individuals, 115 women and 97 men had been evacuated to Sweden or Denmark during the World War II according to the Finnish National Archives' register. To study the effects of early separation from parents on methylome, we selected 83 evacuated men with data on age at and length of evacuation and of father's occupational status and 83 non-evacuated controls matched for sex, birth year and father's occupational status in childhood for the methylation typing. In this study, the analyses are based on 62 evacuated men and 60 non-evacuated controls with full data on CESD depressive symptoms, covariates, and methylation profiles. For this group the mean age was 63.5 years (SD = 2.8).

DNA methylation sample, measurement, normalization and quality control. DNA methylation analysis was performed at the Genetics Core of the Wellcome Trust Clinical Research Facility (Edinburgh, UK). Bisulphite conversion of 500 ng input DNA was carried out using the EZ DNA Methylation Kit (Zymo Research, Freiburg, Germany). Four microlitres of bisulphite-converted DNA was processed using the Infinium HD Assay for Methylation. This was performed using the Illumina Methylation 450 k beadchip and Infinium chemistry (Illumina, Inc., San Diego, CA, USA). Each sample was interrogated on the arrays against 485 000 methylation sites. The arrays were imaged on the Illumina HiScan platform and genotypes were called automatically using GenomeStudio Analysis software version 2011.1.
Quality control pipeline was set up using the R-package minfi, including intensity read outs, normalization, cell type composition estimation, β- and M-value calculation. We excluded samples with low intensity (badSampleCutoff < 10.6) or deviant beta distribution based on visual inspection (n = 5). We did not detect any gender discrepancy. Data were normalized with functional normalization (FunNorm). Of the probes, we excluded those with detection p-value > 0.01 in > 50 % of samples(11), non-autosomal and non-specific binding probes as well as probes with SNPs in the interval for which the Illumina probe is designed to hybridize and if they were located close (10bp from query site) to a SNP which had a minor allele frequency of ≥0.05. Probes located in the X and Y chromosome were also excluded. This yielded a total number of probes of 424,844. Batch effects were identified by inspecting the association of principal

components of the methylation levels with possible technical batches using linear regressions and visual inspection of PCA plots using the Bioconductor R package *shinyMethyl* (version 0.99.3). Identified batch effects (i.e. array column) were removed using the Empirical Bayes' (EB) method *ComBat(12)*. Batch corrected β-values after *ComBat* were used for all further statistical analyses. DNA methylation data from the HBCS is available on request to Jari Lahti e-mail: jari.lahti@helsinki.fi.

Depressive symptoms assessment. Frequency of depressive symptoms were self-reported with the 20-item Centre for Epidemiologic Studies Depression scale (CESD).

Covariates. Age, sex, current smoking status, evacuation status (evacuated in childhood or not), and antidepressant medication (Model 2) were self-reported at the clinical visit. Blood cell proportions were estimated with the Houseman method. First three genotype MDS components were derived from genome-wide genotype data.

Analysis. A linear regression of the CESD depressive symptoms score and covariates on each individual methylation probe in the array was performed in R.

Acknowledgements. We thank all study participants as well as all involved in the Helsinki Birth Cohort Study. Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland (1284859 ja 12848591), the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Signe and Ane Gyllenberg Foundation, University of Helsinki, Ministry of Education, Ahokas Foundation, Emil Aaltonen Foundation, and Scottish Senior Clinical Fellowship (SCD/09). We also acknowledge the Genetics Core of the Wellcome Trust Clinical Research Facility (Edinburgh, UK) who ran the 450k array for the HBCS samples.

Cooperative Health Research in the Augsburg Region (KORA)

Description. The Cooperative Health Research in the Region of Augsburg study is an independent population-based cohort from the region of Augsburg, Southern Germany. Whole blood samples were obtained from the KORA F4 survey (examination 2006-2008), a seven-year follow-up study of the KORA S4 cohort. Participants gave written informed consent and the study was approved by the local ethics committee (Bayerische Landesärztekammer).

DNA methylation sample, measurement, normalization and quality control. Whole blood genomic DNA was bisulfite converted using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA) according to the manufacturer's procedure, with the alternative incubation conditions recommended when using the Illumina Infinium Methylation Assay. Raw methylation data were extracted using the Illumina GenomeStudio software (version 2011.1, Methylation module 1.9.0). Preprocessing was performed using R (version 3.0.1). Probes with signals from less than three functional beads, and probes with a detection p-value > 0.01 were defined as low-confidence probes. Probes that covered SNPs (MAF in Europeans >5%) were excluded from the data set. A color bias adjustment was performed using the R package lumi (version 2.12.0) through smooth quantile normalization and background correction based on the negative control probes present on the Infinium HumanMethylation BeadChip. This was performed separately for the two color channels and chips. β-values corresponding to low-confidence probes were set to missing. A 95% call rate threshold was applied on samples and CpG sites. Beta-mixture quantile normalization (BMIQ) was applied using the R package wateRmelon, version 1.0.3. Because KORA F4 samples were processed on 20 96-well plates in 9 batches, plate and batch effects were investigated using principal component analysis and eigenR2 analysis. The DNA methylation dataset from the KORA study is available upon request from KORA-gen (https://epi.helmholtz-muenchen.de/). Data requests can be submitted online and are subject to approval by the KORA Board.

Depressive symptoms assessment. Depressive symptoms were assessed using the German version of the self-administered Patient Health Questionnaire (PHQ-9).(13) PHQ-9 represents a depression module that scores each of the nine Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV)

criteria for depression from zero to three. PHQ-9 scores of 5, 10, 15, and 20 represents mild, moderate, moderately severe and severe depression. It is used to monitor the severity of the depressive symptoms. PHQ scores higher or equal to 10 have been shown to have a sensitivity of 88% and a specificity of 88% for major depression. PHQ-9 is recommended by the American Psychiatric Association (APA) working group for DSM-5 as instrument to evaluate the severity of major depressive disorder according to the new DSM-5 criteria.

Covariates. The association between DNA methylation level as the outcome and depression as the independent variable was performed using linear regression models, adjusting for age, sex, smoking, antidepressants, white blood cell count, and technical covariates (analytic plate and chip position on plate).
Information on sociodemographic variables including sex, age, smoking status and use of antidepressants was collected by trained medical staff during a standardized interview. The estimated white blood cell proportions were obtained using the method by Houseman et al.(5)

Acknowledgements. The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research has been supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement [n°313010] (Large-scale prospective cohort studies - BBMRI-LPC; www.bbmri-lpc.org), [n°602736] (Multi-dimensional omics approach to stratification of patients with low back pain - PAIN-OMICS; http://www.painomics.eu/) and under grant agreements [n°603288] (Systems Biology to Identify Molecular Targets for Vascular Disease Treatment – SysVasc; http://www.sysvasc.eu/).
The authors are grateful for the commitment and involvement of all the study participants and for the work and dedication of the MONICA/KORA Augsburg Study staff. We thank Nadine Lindemann and Franziska Scharl for technical support in the DNA methylation measurements that were performed at the Helmholtz Zentrum München.

Lothian Birth Cohorts of 1921 (LBC1921) and 1936 (LBC1936)

Description. The LBC1921 and LBC1936 represent relatively healthy older individuals, mostly living in the Lothian region of Scotland, born in 1921 and 1936, respectively. They have been involved in longitudinal assessments of psychological and medical traits from the age of 79 (LBC1921) or 70 (LBC1936) years.(14, 15) Here, we report the first wave, with the largest sample size: 433 (60.3% female) participants with a mean age of 79.1±0.6 years (range = 77-80) in the LBC1921; 920 (49.5% female) participants with a mean age of 69.6±0.8 years (range = 67-71) in the LBC1936.

DNA methylation sample, measurement, normalization and quality control. 514 whole blood samples in LBC1921 and 1,004 samples in LBC1936 were taken and DNA extracted using standard methods at, respectively, MRC Technology, Western General Hospital, Edinburgh and the Wellcome Trust Clinical Research Facility (WTCRF), Western General Hospital, Edinburgh. 485,512 methylation probes were typed at the WTCRF. The R minfi package was used to background correct raw intensity data and to generate methylation beta-values.(2) Probes with a low (<95%) detection rate (at $p < 0.01$) were removed and manual quality control via inspection of the array control probe signals was used to locate and exclude low quality samples (e.g. samples with inadequate hybridisation, bisulfite conversion, nucleotide extension or staining signal). Samples with a low call rate (samples with <450,000 probes detected at $p < 0.01$) were excluded based on the Illumina-recommended thresholds. Because the LBC samples had previously been genotyped (Illumina 610-Quadv1), genotypes derived from the 65 SNP control probes on the methylation array using the wateRmelon package(16) were compared to those on the genotyping array. Exclusions were made for samples with a low match of genotypes with SNP control probes, potentially indicative of sample contamination or mix-up (n = 9). Additionally, 8 subjects whose predicted sex, based on XY probes, did not align with reported sex were eliminated. The DNA methylation

dataset from the Lothian Birth Cohorts is available on requests to the Study Director, Ian Deary e-mail: ian.deary@ed.ac.uk.

Depressive symptoms assessment. Depressive symptoms were assessed with the Hospital Anxiety Depression Scale (17), containing seven items each for anxiety and depression. The total score of the seven depression items was used.
Covariates. Self-reported smoking behavior (current, former and never) and medication use was also ascertained in the clinical testing session. The Anatomical Therapeutical Chemical-Code was used to code anti-depressants for current- and non-users. Age, sex, the first 4 principal components obtained from genotyping array data to index population substructure, B cells, monocytes, CD8T cells, CD4T cells, NK cells, position on BeadChip, methylation sample plate, BeadChip, hybridization date (and smoking for model 2). White blood cell counts were imputed from the methylation array using the Houseman method.(5)

Analysis. A linear regression of the HADS depression score and covariates on each individual methylation probe in the array was performed in R (v3.2.3).

Acknowledgements. We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society, and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Methylation typing was supported by the Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. Analysis of these data were carried out within the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE). CCACE is supported by funding from the BBSRC, the Medical Research Council (MRC), and the University of Edinburgh as part of the cross-council Lifelong Health and Wellbeing initiative (MR/K026992/1). Funding from the Australian NHMRC (613608) supported quality control analysis of the methylation data.

The Rotterdam Study (RS-III & RS-BIOS)

Description. The Rotterdam Study is a prospective population based cohort study in a well-defined suburb in the city of Rotterdam, the Netherlands. The design and prospective of the Rotterdam Study has been described in details elsewhere.(18) For the current analysis we used data from individuals aged 45 years and older that participated in the third cohort of the Rotterdam Study. Samples were obtained in two stages. First stage included participants of the first visit of the third cohort (RS-III-1). Second stage was obtained as a part of the BBMRI-NL (Biobanking and Biomolecular Research Infrastructure Netherlands) – BIOS project and included participants of the fifth visit of the first cohort (RS-I-5), third visit of the second cohort (RS-II-3) and second visit of the third cohort (RS-III-2). The subsets from the two stages were analyzed individually and then meta-analyzed, together with other cohorts.

DNA methylation sample, measurement, normalization and quality control. In both subsets of in total 1479 individuals a whole blood DNA methylation was quantified utilizing same methods. DNA was extracted from whole peripheral blood, stored in EDTA tubes, by standardized salting out methods. The genome-wide DNA-methylation levels were determined using the Illumina HumanMethylation 450K array (Illumina, Inc., San Diego, CA, USA).(19) In short, samples (500ng of DNA per sample) were first treated with bisulfite using the Zymo EZ-96 DNA-methylation kit (Zymo Research, Irvine, CA, USA). Next, in accordance with the manufacturers' protocol the samples were hybridized to the arrays. The methylation percentage of a CpG site was reported as a β-value ranging between 0 for no methylation and 1 for full methylation.
Quality control (QC) of the samples was carried out using Illumina Genome Studio software (v2011.1, methylation module version 1.9.0; Illumina). During QC samples with incomplete bisulfite treatment, with a low detection rate (< 99 %), or gender swaps, were excluded. Also, probes with a detection p-value >

0.01 in > 1 % samples, were filtered out. In total of 474,528 probes passed the QC in the first subset and 419,937 in the BIOS subset (excluding sex chromosome).Filtered β values were than normalized with DASEN implemented in the *wateRmelon* package of the R statistical software.(16) The DNA methylation dataset from the RS should be directed towards the management team of the Rotterdam Study email: secretariat.epi@erasmusmc.nl.

Depressive symptoms assessment. Depressive symptoms were assessed with the Dutch version of the Centre for Epidemiologic Studies Depression scale (CESD).(20) The CESD scale was designed to assess presence and severity of self-reported depressive symptoms.(21) Participants were asked 20 questions that correspond with criterion based-symptoms associated with depression, and they could score from 0-60. The screening for depressive symptoms was performed during the home interview by trained research assistants.

Covariates. Age, sex, current smoking status and antidepressant medication were measured using standard cohort specific protocols at the time DNA samples were collected. During this interview participants reported on their smoking behavior (current, former and never) and psychotropic medication use. Moreover, to confirm the self-reported use of antidepressants prescription, the Anatomical Therapeutical Chemical-Code(22) data was collected from pharmacies linked records. Exposure to an antidepressant was defined as current user and non-user of any antidepressant as described in detail elsewhere.(22) White blood cells counts (monocytes, granulocytes and lymphocytes) were measured immediately at the research center using a standard hematology analyzer (Beckman Coulter, Pasadena, CA, USA).

Analysis. An epigenome-wide linear regression was performed per DNA methylation probe with CESD depression score adjusting for covariates in R. Model 1 was adjusted for age, sex, smoking status, batch effects and cell composition, while Model 2 was additionally adjusted for antidepressant medication use.

Acknowledgements. The generation and management of the Illumina 450K methylation array data (EWAS data) for the Rotterdam Study were executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and the Netherlands. The EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk, and Lisette Stolk PhD for their help in creating the methylation database. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

Generation Scotland Study (GS)

Description. Generation Scotland: Scottish Family Health Study (GS:SFHS) is a population- and family-based cohort comprising ~24,000 individuals from the Scottish population. The cohort has been described in detail previously.(23) Two hundred and twenty-seven individuals (118 cases, 109 controls) were included in the current study. These individuals are of European descent and were aged 40 years or over at the time of recruitment to GS:SFHS. Antidepressant usage was determined by self-report at the time of recruitment. Blood samples for DNA extraction were obtained at the time of recruitment to GS:SFHS.

DNA methylation sample, measurement, normalization and quality control. Whole blood genomic DNA (500 ng) was treated with sodium bisulphite using the EZ-96 DNA Methylation Kit (Zymo Research, Irvine, California), according to the manufacturer's instructions. DNA methylation was assessed using the

Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, California), according to the manufacturers protocol. Samples were assigned to slides such that gender was counter-balanced. Methylation array processing was carried out at the Wellcome Trust Clinical Research Facility, The University of Edinburgh.
Raw intensity (.idat) files were read into R using the minfi package(2), which was used to perform initial quality control assessments by assessing signal from the array's internal control probes. All samples were deemed to have performed well based on the output of the control probes.
Probe and sample filtering was performed using the pfilter function in wateRmelon.(16) Probes were excluded if they had a detection *p*-value ≥ 0.05 in more than 5% samples (n = 1000) or had a beadcount of less than 3 in more than 5% samples (n = 466). All samples satisfied an inclusion criterion of having no more than 5% probes with a detection *p*-value ≥ 0.01. The final dataset comprised 484145 probes measured in 227 individuals.
The raw data (β-values) were normalised using the dasen function in wateRmelon.(16) The DNA methylation data from the Generation Scotland (GS) data is available on request to e-mail: access@generationscotland.org.
GS has Research Tissue Bank status, and the GS Access Committee reviews applications to ensure that they comply with legal requirements, ethics and patient consent.

Assessment of depression. Diagnoses of depression were established by trained researchers who administered the screening questions of the Structured Clinical Interview for DSM-IV (SCID)(24) to all GS:SFHS participants. Those who screened positive were subsequently assessed using the mood disorder sub-sections of the SCID. Major depressive disorder (MDD) was diagnosed in those meeting the criteria for MDD whose symptoms could not be better explained by bipolar disorder, a general medical condition or substance abuse.

Covariates. The following variables were included as covariates: sex, age, current smoking status; family; estimated blood cell counts; batch (methylation slide ID); and methylation-based principal components. Antidepressant usage was additionally included as a covariate in a sensitivity analysis carried out to assess the contribution of antidepressant use to depression-associated methylation differences.
Smoking status was determined by self-report at the time of blood sample collection and was recorded as a binary variable (current smokers vs. non-smokers (never smokers and ex-smokers)). Antidepressant usage was also determined by self-report at the time of blood sample collection. No data on antidepressant usage was available for seven individuals.
Estimated proportions of six blood cell types (monocytes, granulocytes, CD4+ T-cells, CD8+ T-cells, B-cells and natural killer cells) were obtained using the estimateCellCounts function in minfi.(5, 25) In order to avoid collinearity, one cell type, natural killer cells, was excluded from the analyses.
Methylation-based principal components were identified using prcomp without scaling. PCA was performed on the matrix of residualised β-values obtained after residualising for the other covariates included in the model used to identify differentially methylated positions (principal components were estimated separately for model 1 and model 2). Principal components that individually accounted for at least 1% of the variance were included as covariates. For both models, this resulted in the inclusion of the first six principal components.
As our sample comprised related individuals (100 discordant sib pairs and 27 singletons), it was necessary to account for relatedness. This was achieved using the duplicateCorrelation function in limma(26, 27) to fit family as a random effect.

Analysis. Differences in DNA methylation associated with the diagnosis of depression were detected using the following linear mixed effects models (family was treated as a random effect and all other covariates as fixed effects), implemented in limma(27):

Model 1 (primary analysis):
Methylation (β-values) ~ Depression diagnosis + Gender + Age + Current smoking status + Estimated blood cell counts (monocytes, granulocytes, CD4+ T-cells, CD8+ T-cells and B-cells) + Batch + PCs 1-6 + Family

Model 2 (sensitivity analysis to assess the effect of co-varying for antidepressant usage): Methylation (β-values) ~ Depression diagnosis + Gender + Age + Current smoking status + Estimated blood cell counts (monocytes, granulocytes, CD4+ T-cells, CD8+ T-cells and B-cells) + Batch + PCs 1-6 + Antidepressants + Family

Acknowledgements. We are grateful to the participants for their participation in this study. KLE would like to acknowledge the support of the Brain & Behavior Research Foundation through a NARSAD Independent Investigator Award, which funded the methylation analysis. DNA extraction, bisulphite conversion and the processing of the Infinium HumanMethylation450 BeadChips were performed at the Wellcome Trust Clinical Research Facility, Edinburgh, UK. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z). RMW received salary support from award no. R01 MH102068-02 from the National Institutes of Health to DJP, R. McCombie and others. KLE, RMW, AMM and DJP are members of The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE), part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding of CCACE from the BBSRC, EPSRC, ESRC and MRC is gratefully acknowledged.

Atherosclerosis Risk In Communities (ARIC) Study

Description. The ARIC study is a population-based cohort study of atherosclerosis and clinical atherosclerotic diseases.(28) At its inception (1987-1989), 15,792 men and women, including 11,478 white and 4,266 black participants were recruited from four U.S. communities: Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina; and Jackson, Mississippi. In the first 3 communities, the sample reflects the demographic composition of the community. In Jackson, only African-American residents were enrolled. Participants included in this project were 2,297 African-Americans with DNA methylation and depressive symptoms measures, both collected at visit 2.
DNA methylation sample, measurement, normalization and quality control. DNA methylation analysis was conducted with the Infinium HumanMethylation450 BeadChip (HM450) array (Illumina Inc., San Diego, CA) on genomic DNA extracted from blood samples collected at ARIC Visit 2. Assay were performed on participants who had not restricted use of their DNA and for whom at least 1 μg of DNA and genome-wide genotyping data were available. Details of assay and QC procedures have been previously published.(29) Briefly, genomic DNA was treated with sodium bisulfite using the EZ-96 DNA methylation kit (Zymo Research Corporation, Irvine, CA) following the manufacturer's protocol. Bisulfite converted DNA was amplified, enzymatically fragmented, purified and hybridized to the HM450 array in accordance with the manufacturer's directions. Methylation typing at 485,577 CpG sites was performed using GenomeStudio 2011.1 (Illumina Inc., San Diego, CA). Methylation level for each probe was derived as a beta value representing the fractional level of methylation at that location. Quality control analysis was performed using the wateRmelon R package.(16) Probe data were excluded if they had a low detection rate (<95% at $p<0.01$) and a high missing rate (greater than 1% across all samples). Sample data were excluded based on the following criteria: (1) greater than 5% missing values across all probes; (2) possible gender mismatch based on principal component analysis; (3) genotype mismatch based on 24 SNPs present on the HM450 array. Methylation values were normalized using the Beta MIxture Quantile dilation (BMIQ) method.(30) All ARIC Study participants have provided written informed consent for genomic studies and this work was conducted in compliance with the Helsinki Declaration. The Committee for the Protection of Human Subjects at the University of Texas Health Science Center at Houston has approved this research. The DNA methylation dataset from the ARIC Study is available on request at https://www2.cscc.unc.edu/aric/distribution-agreements.

Depressive symptoms assessment. Depressive symptoms were assessed at visit 2 using the 21-item Maastricht Questionnaire.(31) Responses to the questionnaire are coded as yes = 2, don't know = 1, and

no = 0. Two items, questions 9 and 14, are reversed coded (yes = 0, don't know = 1, no = 2). Responses are summed to obtain an overall vital exhaustion score, which ranges from 0 to 42, with higher scores representing more exhaustion. Cronbach alpha for internal consistency has been reported as 0.89.(31) Although designed to measure vital exhaustion, the 21-item Maastricht Questionnaire has been shown to correlate with measures of depressive symptoms(32) and was previously used to assess depressive symptoms.(33, 34)

Covariates. Age, sex, current smoking status and antidepressant medication were measured using standard cohort specific protocols at visit 2. Blood cell proportions were imputed using the Houseman method.(5) Genotype principal components were derived from genome-wide genotype data (Affymetrix 6.0).

Analysis. Analysis was performed according to a pre-specified analysis plan adjusting for biological and technical covariates (see attached spreadsheet for details).

Acknowledgments. The Atherosclerosis Risk in Communities (ARIC) study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for “Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419).

Women’s Health Initiative - Epigenetic Mechanisms of PM-Mediated CVD (WHI-EMPC)

Description. WHI-EMPC is an ancillary study of epigenetic mechanisms underlying associations between ambient particulate matter (PM) air pollution and cardiovascular disease (CVD) in the Women’s Health Initiative clinical trials (CT) cohort. It is funded by the National Institute of Environmental Health Sciences (R01-ES020836).(35)
The WHI-EMPC study population is a stratified, random sample of 2,200 WHI CT participants who were examined between 1993 and 2001; had available buffy coat, core analytes, electrocardiograms, and ambient concentrations of PM; but were not taking anti-arrhythmic medications at the time.
As such, WHI-EMPC is representative of the larger, multiethnic WHI CT population from which it was sampled: n = 68,132 participants aged 50-79 years who were randomized to hormone therapy, calcium / vitamin D supplementation, and / or dietary modification in 40 U.S. clinical centers at the baseline exam (1993-1998) and re-examined in the fasting state one, three, six, and nine years later.(36)
During participant visits, data on age, race/ethnicity, smoking status, depressive symptoms, and antidepressant medication use were obtained. Current analyses were in European Americans and involved information collected at the first visit with available DNA methylation (DNAm) data.

DNA Methylation. Genome-wide DNAm at CpG sites was measured using the Illumina 450K Infinium Methylation BeadChip, quantitatively represented by beta (the percentage of methylated cytosines over the sum of methylated and unmethylated cytosines), and quality controlled using the following filters: detection p-values > 0.01 in > 10% of samples, detection p-values > 0.01 in > 1% of probes, yielding values of beta at 484,220 sites. DNAm data was normalized using BMIQ(30), then stage- and plate-adjusted using ComBat. [12] Modeled epigenome-wide associations also adjusted for principal components for ancestral admixture, cell subtype proportions (CD8-T, CD4-T, B cell, natural killer, monocyte, and granulocyte)(5), and technical covariates including array, row, and column. Researchers seeking opportunities to access Women's Health Initiative data are encouraged to solicit collaboration with a WHI Principal Investigator who can sponsor research proposals and provide expertise on study design, content, and data analysis. WHI Principal Investigators are listed at https://www.whi.org/researchers/SitePages/WHI%20Investigators.aspx. Standing policies governing

research proposals, data access, data use agreements, presentations and publications are available at https://www.whi.org/researchers/SitePages/Write%20a%20Paper.aspx.

Depressive symptoms ascertainment. Depressive symptoms were obtained on the day of DNAm collection using a short screening instrument developed by Burnman et al.(37) that combines six items from the CES-D that highly correlate with the full instrument (r =0.88)(38) with two items from the Diagnostic Interview Schedule (DIS). This CES-D/DIS screening instrument ranges from 0 to 1 with a higher score indicating a greater likelihood of depression.

Acknowledgments. The Women's Health Initiative Epigenetic Mechanisms of PM-Mediated CVD (WHI-EMPC) is funded by the the National Institute of Environmental Health Sciences (R01-ES020836; Whitsel, Baccarelli, Hou). The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

REFERENCES

1. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. (1991): The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1:263-276.
2. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. (2014): Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 30:1363-1369.
3. Maksimovic J, Gordon L, Oshlack A (2012): SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol*. 13:R44.
4. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. (2014): Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol*. 15:503.
5. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. (2012): DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 13:86.
6. Dawber TR, Meadors GF, Moore FE, Jr. (1951): Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health*. 41:279-281.
7. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP (1979): An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 110:281-290.
8. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. (2013): Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 8:203-209.
9. Radloff LS (1977): The CES-D scale: a self report depression scale for research in general population . *Appl Psychol Measurement*.385-401.
10. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP (1975): The Framingham Offspring Study. Design and preliminary data. *Prev Med*. 4:518-525.
11. Amin N, Jovanova O, Adams HH, Dehghan A, Kavousi M, Vernooij MW, et al. (2016): Exome-sequencing in a large population-based study reveals a rare Asn396Ser variant in the LIPG gene associated with depressive symptoms. *Mol Psychiatry*.
12. Johnson WE, Li C, Rabinovic A (2007): Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 8:118-127.
13. Kroenke K, Spitzer RL, Williams JB (2001): The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 16:606-613.
14. Luciano M, Houlihan LM, Harris SE, Gow AJ, Hayward C, Starr JM, et al. (2010): Association of existing and new candidate genes for anxiety, depression and personality traits in older people. *Behav Genet*. 40:518-532.
15. Deary IJ, Gow AJ, Pattie A, Starr JM (2012): Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 41:1576-1584.
16. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC (2013): A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 14:293.
17. Zigmond AS, Snaith RP (1983): The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 67:361-370.
18. Hofman A, Brusselle GG, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, et al. (2015): The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol*. 30:661-708.
19. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. (2011): Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*. 6:692-702.
20. Breslau N (1985): Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES-D and results from diagnostic interviews. *Psychiatry Res*. 15:219-229.

21. Hales DP, Dishman RK, Motl RW, Addy CL, Pfeiffer KA, Pate RR (2006): Factorial validity and invariance of the center for epidemiologic studies depression (CES-D) scale in a sample of black and white adolescent girls. *Ethn Dis*. 16:1-8.
22. Noordam R, Aarts N, Verhamme KM, Sturkenboom MC, Stricker BH, Visser LE (2015): Prescription and indication trends of antidepressant drugs in the Netherlands between 1996 and 2012: a dynamic population-based study. *Eur J Clin Pharmacol*. 71:369-375.
23. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. (2013): Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol*. 42:689-700.
24. First M, Gibbon, M., Spitzer, R., and Willaims, J. Structured clinical interview for DSM-IV-TR Axis Disorder (Research Version). SCID-I/P, 2002.
25. Jaffe AE, Irizarry RA (2014): Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 15:R31.
26. Smyth GK, Michaud J, Scott HS (2005): Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*. 21:2067-2075.
27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. (2015): limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 43:e47.
28. (1989): The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 129:687-702.
29. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, et al. (2015): Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet*. 24:4464-4479.
30. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. (2013): A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 29:189-196.
31. Appels A, Hoppener P, Mulder P (1987): A questionnaire to assess premonitory symptoms of myocardial infarction. *Int J Cardiol*. 17:15-24.
32. Kopp MS, Falger PR, Appels A, Szedmak S (1998): Depressive symptomatology and vital exhaustion are differentially related to behavioral risk factors for coronary artery disease. *Psychosom Med*. 60:752-758.
33. Wojciechowski FL, Strik JJ, Falger P, Lousberg R, Honig A (2000): The relationship between depressive and vital exhaustion symptomatology post-myocardial infarction. *Acta Psychiatr Scand*. 102:359-365.
34. Wattanakit K, Williams JE, Schreiner PJ, Hirsch AT, Folsom AR (2005): Association of anger proneness, depression and low social support with peripheral arterial disease: the Atherosclerosis Risk in Communities Study. *Vasc Med*. 10:199-206.
35. Goltser-Dubner T, Galili-Weisstub E, Segman RH (2010): Genetics of unipolar major depressive disorder. *Isr J Psychiatry Relat Sci*. 47:72-82.
36. (1998): Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials*. 19:61-109.
37. Burnam MA, Wells KB, Leake B, Landsverk J (1988): Development of a brief screening instrument for detecting depressive disorders. *Med Care*. 26:775-789.
38. Wassertheil-Smoller S, Shumaker S, Ockene J, Talavera GA, Greenland P, Cochrane B, et al. (2004): Depression and cardiovascular sequelae in postmenopausal women. The Women's Health Initiative (WHI). *Arch Intern Med*. 164:289-298.

eTable 1. Top DNA methylation sites associated with depressive symptoms in the discovery EWAS: M2 is model additionally adjusted for antidepressant medication.

CpG site ID	Chr	Location	Gene symbol	P-value M1	P-value M2
cg04987734	14	103415873	*CDC42BPB*	4.93×10^{-8}	3.94×10^{-8}
cg07012687	17	80195180	*SLC16A3*	3.47×10^{-7}	4.51×10^{-7}
cg08796240	16	70733832	*VAC14*	7.43×10^{-7}	4.06×10^{-6}
cg06096336	2	231989800	*PSMD1;HTR2B*	8.06×10^{-7}	2.11×10^{-6}
cg16745930	10	100220809	*HPSE2*	1.34×10^{-6}	2.19×10^{-6}
cg09849319	5	1494983	*LPCATI1*	1.81×10^{-6}	6.14×10^{-6}
cg17237086	22	40814966	*MKL1*	3.44×10^{-6}	1.30×10^{-5}
cg03985718	2	105924245	*TGFBRAP1*	3.61×10^{-6}	1.28×10^{-5}
cg21098005	20	44538605	*PLTP*	4.36×10^{-6}	7.86×10^{-6}
cg16466652	19	6271960	*MLLT1*	4.39×10^{-6}	7.08×10^{-6}
cg07884764	11	64107517	*CCDC88B*	5.03×10^{-6}	1.50×10^{-5}
cg01541347	7	4729920	*FOXK1*	5.64×10^{-6}	6.66×10^{-6}
cg02341197	21	34185927	*C21or62*	5.84×10^{-6}	8.57×10^{-6}
cg01947751	3	196728969	-	6.23×10^{-6}	5.93×10^{-6}
cg13747876	17	80195402	*SLC16A3*	6.32×10^{-6}	1.31×10^{-5}
cg12764201	1	105101123	*CORT;APITD1*	7.15×10^{-6}	1.33×10^{-5}
cg08295111	5	133866097	*PHF15*	7.87×10^{-6}	1.73×10^{-5}
cg18030453	3	45506216	*LARS2*	9.16×10^{-6}	2.44×10^{-5}
cg12325605	3	56810151	*ARHGEF3*	9.62×10^{-6}	1.82×10^{-5}
cg23282441	10	73533927	*C10orf54;CDH23*	9.69×10^{-6}	5.30×10^{-6}

eTable 2. Top DNA methylation sites associated with depressive symptoms in the meta-EWAS (N = 11,256).

MarkerName	Weight	Zscore	P.value	Direction[a]	HetPVal	CHR	MAPINFO[b]	UCSC_RefGene
cg12325605	11256	5.84	**5.24×10^{-09}**	++++++++++	0.90	3	56810151	*ARHGEF3*
cg04987734	11256	5.65	**1.57×10^{-08}**	++--+-++++-	4.87×10^{-03}	14	103415873	*CDC42BPB*
cg14023999	11256	5.42	**5.99×10^{-08}**	+++++++++-++	0.15	15	90543224	
cg18030453	11253	5.29	1.20×10^{-07}	+++-+-++++-	0.47	3	45506216	*LARS2*
cg10829227	11254	5.10	3.40×10^{-07}	+++++-+++++	0.57	19	47200595	*PRKD2*
cg07175797	11254	5.07	3.92×10^{-07}	+++-+++-+++	0.39	16	50317656	
cg07012687	11251	5.05	4.46×10^{-07}	+++-++++++-	0.27	17	80195180	*SLC16A3*
cg12728588	11256	5.01	5.43×10^{-07}	++++++++++	0.58	1	36025489	*NCDN*
cg22069247	11253	4.93	8.23×10^{-07}	+++-+-+++++	0.74	2	232393256	*NMUR1*
cg12526091	11256	4.91	9.00×10^{-07}	+++-+-+++++	0.85	12	58245042	
cg14012686	11253	4.89	1.01×10^{-06}	++-+++++++-	0.62	2	74785750	*C2orf65*
cg00153395	11255	4.89	1.02×10^{-06}	+++-+++++++	0.84	1	65327523	*JAK1*
cg04583842	10491	4.84	1.33×10^{-06}	+++-+++?+++	0.31	16	88103117	*BANP*
cg03429645	11255	4.82	1.43×10^{-06}	+++++++-+++	0.67	3	100053188	*NIT2*
cg04286697	11256	4.78	1.75×10^{-06}	+++-+++++++	0.29	2	232259623	*B3GNT7*
cg08796240	11256	4.78	1.80×10^{-06}	++++++++++-	0.25	16	70733832	*VAC14*
cg19769147	11255	4.74	2.09×10^{-06}	+++++-++++-	0.17	14	105860954	*PACS2*
cg17822325	11253	4.72	2.33×10^{-06}	+++++-+++++	0.93	1	31896462	*SERINC2*
cg06096336	11254	4.71	2.51×10^{-06}	++-++-+++++	0.43	2	231989800	*PSMD1; HTR2B*
cg10401362	11254	4.69	2.71×10^{-06}	-++-+-+++++	1.20×10^{-03}	7	157185402	*DNAJB6*
cg07467649	11256	4.69	2.75×10^{-06}	+++-+-+++++	0.36	7	72856462	*BAZ1B*
cg13747876	11256	4.68	2.93×10^{-06}	+++++-++++-	0.49	17	80195402	*SLC16A3*
cg25392060	11256	4.66	3.10×10^{-06}	+++-+++++++	0.52	8	142297121	
cg22365240	11250	4.66	3.17×10^{-06}	-++++++++++	0.98	2	105374995	
cg08631783	11253	4.62	3.83×10^{-06}	-++-+-+-+++	0.13	15	89418456	*ACAN*
cg19269039	11255	4.62	3.84×10^{-06}	+++++++++-++	0.51	1	111743200	*DENND2D*
cg03720762	11256	4.57	4.82×10^{-06}	+++++++-+++	0.09	17	17604184	*RAI1*
cg21604136	11255	4.57	4.93×10^{-06}	++-++++++++	0.20	1	9910137	*CTNNBIP1*
cg11931558	11252	4.55	5.46×10^{-06}	++++++++++	0.82	7	11013742	*PHF14*
cg24550880	11255	4.53	5.88×10^{-06}	+++-+-+++++	0.46	17	79420279	*BAHCC1*
cg07372520	11255	4.53	6.01×10^{-06}	+++-+-+++++	0.42	1	180086434	
cg05827190	11255	4.52	6.08×10^{-06}	+++++-++-++	0.23	4	681440	*MFSD7*
cg19743103	11253	4.52	6.08×10^{-06}	+++--++++++	0.14	1	6304220	*HES3*
cg17237086	11256	4.52	6.09×10^{-06}	+++++-++++-	0.35	22	40814966	*MKL1*
cg16745930	11254	-4.52	6.26×10^{-06}	----------	0.11	10	100220809	*HPSE2*
cg27305772	11256	4.51	6.57×10^{-06}	+++++-+++++	0.83	11	65630355	*MUS81*
cg02341197	11256	4.50	6.80×10^{-06}	+++++-+++++	0.86	21	34185927	*C21orf62*
cg23606718	11255	4.46	8.37×10^{-06}	+++-+++++++	0.81	2	131513927	*FAM123C*
cg23282441	11254	4.45	8.63×10^{-06}	++--++++++-	0.43	10	73533927	*C10orf54; CDH23*
cg17743381	11256	4.44	9.10×10^{-06}	+++++-+++++	0.70	1	39024825	
cg05251389	11254	4.42	9.67×10^{-06}	+++-+++-+++	0.71	22	43525330	*BIK*
cg26610247	11255	4.42	9.70×10^{-06}	++---++++++	0.17	8	142297175	

[a] Direction of association in all cohorts. Order of cohorts in the direction column: CHS, FHS, GS, HBCS, KORA, LBC1921, LBC1936, RS_BIOS, RSIII, ARIC and WHI-EMPC; [b] Genomic position based on human genome 19 (hg19)

eTable 3: Cohort Specific association results for the top three methylation sites.

			cg04987734			cg12325605			cg14023999		
Cohort	N	Instrument	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
ARIC	2297	21-MQ	0.0004	0.0001	**2.27E-03**	0.0003	0.0001	**2.10E-03**	0.0004	0.0001	**6.39E-06**
CHS	323	CESD(10 item)	0.0007	0.0006	2.79E-01	0.0003	0.0005	4.50E-01	0.0019	0.0006	**1.40E-03**
FHS	2722	CESD(20 item)	0.0005	0.0001	**5.98E-04**	0.0003	0.0001	**3.11E-03**	0.0003	0.0002	**4.44E-02**
GSncAD	227	SCID	-0.0010	0.0042	8.12E-01	0.0002	0.0029	9.45E-01	0.0037	0.0037	3.15E-01
HBCS	122	CESD(20 item)	-0.0010	0.0007	1.94E-01	0.0002	0.0007	7.67E-01	0.0008	0.0008	3.49E-01
KORA	1727	PHQ-9	0.0138	0.0061	**2.47E-02**	0.0063	0.0047	1.80E-01	0.0055	0.0046	2.30E-01
LBC1921	432	HADS	-0.0015	0.0012	2.31E-01	0.0004	0.0013	7.59E-01	0.0008	0.0013	5.44E-01
LBC1936	916	HADS	0.0025	0.0007	**6.06E-04**	0.0009	0.0008	2.73E-01	0.0020	0.0010	**3.99E-02**
RS-BIOS	757	CESD(20 item)	0.0004	0.0002	1.30E-01	0.0004	0.0002	**2.90E-02**	0.0003	0.0003	2.35E-01
RSIII	722	CESD(20 item)	0.0009	0.0003	**5.61E-04**	0.0006	0.0002	**1.21E-02**	0.0000	0.0003	9.84E-01
WHI	1011	CESD-DIS	-0.0013	0.0013	3.06E-01	0.0026	0.0011	**1.45E-02**	0.0006	0.0013	6.30E-01

eTable 4: Significantly associated gene expression (*cis*) probes with the top CpG sites in blood

PValue	SNPName	SNPChr	SNPPos	ProbeName	ProbeChr	ProbePos	CisTrans	SNPType	AlleleAssessed	OverallZScore	Total N	HGNCName	FDR
1.81E-05	cg14023999	15	90543224	ENSG00000185033	15	90703836	cis	C/T	C	-4.2877145	2101	SEMA4B	0.00465427
2.45E-06	cg04987734	14	103415825	ENSG00000198752	14	103523799	cis	C/T	C	4.7123758	2101	CDC42BPB	7.77E-04

eTable 5: Results of tissue-specific gene expression association with major depressive disorder, inflammation and smoking

Tissue	Trait	gene_name	zscore	effect_size	pvalue	var_g	pred_perf_r2	pred_perf_pval	pred_perf_qval	n_snps_used	n_snps_in_cov	n_snps_in_model
TW_Adipose_Visceral_Omentum_0.5.db.CPD.csv:ENSG00000163947.7	Smoking	ARHGEF3	0.573826	0.087254	0.566085	0.111695	0.089826	3.40E-05	7.19E-05	29	35	35
TW_Cells_Transformed_fibroblasts_0.5.db.CPD.csv:ENSG00000163947.7	Smoking	ARHGEF3	-1.40383	-0.86543	0.16037	0.010835	0.018025	0.026825161	0.010336124	10	12	12
TW_Colon_Sigmoid_0.5.db.CPD.csv:ENSG00000163947.7	Smoking	ARHGEF3	1.761678	0.446355	0.078124	0.055562	0.066145	0.003931831	0.006492176	25	27	27
TW_Esophagus_Mucosa_0.5.db.CPD.csv:ENSG00000163947.7	Smoking	ARHGEF3	-0.65451	-0.16739	0.512785	0.070505	0.074162	1.81E-05	1.53E-05	38	41	41
TW_Pancreas_0.5.db.CPD.csv:ENSG00000163947.7	Smoking	ARHGEF3	1.67459	0.437689	0.094015	0.058995	0.038509	0.016459269	0.014848021	24	30	30
TW_Pituitary_0.5.db.CPD.csv:ENSG00000163947.7	Smoking	ARHGEF3	-1.72174	-0.48817	0.085116	0.042863	0.068916	0.014032363	0.022131184	7	7	7
TW_Skin_Not_Sun_Exposed_Suprapubic_0.5.db.CPD.csv:ENSG00000163947.7	Smoking	ARHGEF3	1.25774	0.368267	0.208486	0.031532	0.028805	0.017399678	0.013340492	14	15	15
TW_Whole_Blood_0.5.db.CPD.csv:ENSG00000163947.7	Smoking	ARHGEF3	1.012799	1.146626	0.311156	0.020622	0.013538	0.032482103	0.015078958	11	11	11
TW_Adipose_Visceral_Omentum_0.5.db.CRP.csv:ENSG00000163947.7	Inflammation	ARHGEF3	-0.29108	-0.00248	0.770994	0.111695	0.089826	3.40E-05	7.19E-05	29	35	35
TW_Cells_Transformed_fibroblasts_0.5.db.CRP.csv:ENSG00000163947.7	Inflammation	ARHGEF3	0.320047	0.010991	0.748933	0.010835	0.018025	0.026825161	0.010336124	10	12	12
TW_Colon_Sigmoid_0.5.db.CRP.csv:ENSG00000163947.7	Inflammation	ARHGEF3	-0.54925	-0.00879	0.582837	0.055562	0.066145	0.003931831	0.006492176	25	27	27
TW_Esophagus_Mucosa_0.5.db.CRP.csv:ENSG00000163947.7	Inflammation	ARHGEF3	-0.26342	-0.0027	0.792227	0.070505	0.074162	1.81E-05	1.53E-05	38	41	41
TW_Pancreas_0.5.db.CRP.csv:ENSG00000163947.7	Inflammation	ARHGEF3	-0.22747	-0.00423	0.820055	0.058995	0.038509	0.016459269	0.014848021	24	30	30
TW_Pituitary_0.5.db.CRP.csv:ENSG00000163947.7	Inflammation	ARHGEF3	0.138821	0.002983	0.889592	0.042863	0.068916	0.014032363	0.022131184	7	7	7
TW_Skin_Not_Sun_Exposed_Suprapubic_0.5.db.CRP.csv:ENSG00000163947.7	Inflammation	ARHGEF3	-1.19529	-0.03317	0.231973	0.031532	0.028805	0.017399678	0.013340492	14	15	15
TW_Whole_Blood_0.5.db.CRP.csv:ENSG00000163947.7	Inflammation	ARHGEF3	0.333525	0.007544	0.738738	0.020622	0.013538	0.032482103	0.015078958	11	11	11
TW_Adipose_Visceral_Omentum_0.5.db.PGC.MDD.csv:ENSG00000163947.7	Major depression	ARHGEF3	0.412344	0.026542	0.680087	0.076475	0.089826	3.40E-05	7.19E-05	21	35	35
TW_Cells_Transformed_fibroblasts_0.5.db.PGC.MDD.csv:ENSG00000163947.7	**Major depression**	**ARHGEF3**	**3.294991**	**0.478392**	**0.000984**	**0.011767**	**0.018025**	**0.026825161**	**0.010336124**	**7**	**12**	**12**
TW_Colon_Sigmoid_0.5.db.PGC.MDD.csv:ENSG00000163947.7	Major depression	ARHGEF3	-0.03845	-0.0048	0.96933	0.037719	0.066145	0.003931831	0.006492176	17	27	27
TW_Esophagus_Mucosa_0.5.db.PGC.MDD.csv:ENSG00000163947.7	Major depression	ARHGEF3	1.307325	0.169943	0.191102	0.014953	0.074162	1.81E-05	1.53E-05	17	41	41
TW_Pancreas_0.5.db.PGC.MDD.csv:ENSG00000163947.7	Major depression	ARHGEF3	-1.31115	-0.15241	0.189806	0.025828	0.038509	0.016459269	0.014848021	18	30	30
TW_Pituitary_0.5.db.PGC.MDD.csv:ENSG00000163947.7	Major depression	ARHGEF3	0.577174	0.062648	0.563822	0.02709	0.068916	0.014032363	0.022131184	5	7	7
TW_Skin_Not_Sun_Exposed_Suprapubic_0.5.db.PGC.MDD.csv:ENSG00000163947.7	Major depression	ARHGEF3	-0.88823	-0.11591	0.374415	0.021688	0.028805	0.017399678	0.013340492	8	15	15
TW_Whole_Blood_0.5.db.PGC.MDD.csv:ENSG00000163947.7	Major depression	ARHGEF3	0.688083	0.092211	0.4914	0.013738	0.013538	0.032482103	0.015078958	5	11	11
TW_Brain_Nucleus_accumbens_basal_ganglia_0.5.db.CPD.csv:ENSG00000198752.5	Smoking	CDC42BPB	-1.13679	-0.17847	0.255628	0.196678	0.071375	0.009631025	0.016237748	40	49	49
TW_Skin_Not_Sun_Exposed_Suprapubic_0.5.db.CPD.csv:ENSG00000198752.5	Smoking	CDC42BPB	0.737376	0.346989	0.460894	0.019126	0.023587	0.031624402	0.022043433	9	15	15
TW_Uterus_0.5.db.CPD.csv:ENSG00000198752.5	Smoking	CDC42BPB	0.696997	0.085148	0.485805	0.185883	0.159989	0.000602084	0.003284848	34	42	42
TW_Brain_Nucleus_accumbens_basal_ganglia_0.5.db.CRP.csv:ENSG00000198752.5	Inflammation	CDC42BPB	-0.94649	-0.00942	0.343897	0.206345	0.071375	0.009631025	0.016237748	41	49	49
TW_Skin_Not_Sun_Exposed_Suprapubic_0.5.db.CRP.csv:ENSG00000198752.5	Inflammation	CDC42BPB	-0.66407	-0.02531	0.506646	0.019126	0.023587	0.031624402	0.022043433	9	15	15
TW_Uterus_0.5.db.CRP.csv:ENSG00000198752.5	Inflammation	CDC42BPB	1.266537	0.013667	0.205321	0.185883	0.159989	0.000602084	0.003284848	34	42	42

TW_Brain_Nucleus_accumbens_basal_ganglia_0.5.db.PGC.MDD.csv:ENSG00000198752.5	**Major depression**	**CDC42BPB**	**-3.00471**	**-0.14497**	**0.002658**	**0.12281**	**0.071375**	**0.009631025**	**0.016237748**	**33**	**49**	**49**
TW_Skin_Not_Sun_Exposed_Suprapubic_0.5.db.PGC.MDD.csv:ENSG00000198752.5	Major depression	CDC42BPB	-0.14192	-0.02278	0.887144	0.015369	0.023587	0.031624402	0.022043433	9	15	15
TW_Uterus_0.5.db.PGC.MDD.csv:ENSG00000198752.5	Major depression	CDC42BPB	2.249692	0.08074	0.024469	0.173171	0.159989	0.000602084	0.003284848	33	42	42
TW_Adipose_Visceral_Omentum_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	0.938201	0.329019	0.348141	0.029282	0.058048	0.000954481	0.001370544	8	9	9
TW_Artery_Aorta_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	2.008999	0.955071	0.044537	0.017738	0.021501	0.039769694	0.022096813	17	23	23
TW_Artery_Tibial_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	1.897563	1.311104	0.057754	0.007106	0.012703	0.057380821	0.0238277	4	6	6
TW_Brain_Putamen_basal_ganglia_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	-1.19224	-0.18487	0.233168	0.208912	0.082556	0.00886192	0.017125428	43	48	48
TW_Cells_EBV-transformed_lymphocytes_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	0.234431	0.175335	0.81465	0.054498	0.035753	0.04392222	0.041637927	21	26	26
TW_Esophagus_Mucosa_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	2.2019	0.591895	0.027672	0.063085	0.145442	9.20E-10	1.43E-09	32	39	39
TW_Heart_Atrial_Appendage_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	0.30938	-0.01417	0.757032	0.020932	0.025039	0.046356302	0.034806914	26	29	29
TW_Heart_Left_Ventricle_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	1.05791	0.130107	0.290096	0.075275	0.141574	8.79E-08	2.59E-07	20	23	23
TW_Lung_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	0.810632	0.074346	0.417577	0.021566	0.024632	0.008760454	0.005973537	18	21	21
TW_Muscle_Skeletal_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	0.421554	0.231807	0.673351	0.0118	0.104255	3.33E-10	5.83E-10	10	13	13
TW_Pancreas_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	2.008163	-0.99542	0.044626	0.066281	0.056007	0.003662688	0.004143342	18	24	24
TW_Pituitary_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	1.162093	0.214899	0.245198	0.323144	0.194396	1.92E-05	0.000105887	63	78	78
TW_Skin_Not_Sun_Exposed_Suprapubic_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	2.256836	0.438941	0.024018	0.081822	0.188317	2.10E-10	8.25E-10	18	23	23
TW_Skin_Sun_Exposed_Lower_leg_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	1.745065	0.235792	0.080974	0.094417	0.171147	6.47E-14	1.40E-13	48	62	62
TW_Spleen_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	1.007384	0.282903	0.31375	0.059964	0.048955	0.037183669	0.036604122	26	28	28
TW_Stomach_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	-0.28295	-0.10299	0.777212	0.044824	0.053373	0.002436732	0.003789078	13	14	14
TW_Whole_Blood_0.5.db.CPD.csv:ENSG00000185033.10	Smoking	SEMA4B	0.842899	0.295704	0.399285	0.088636	0.121142	4.64E-11	7.68E-11	29	33	33
TW_Adipose_Visceral_Omentum_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	0.171927	0.004716	0.863495	0.029282	0.058048	0.000954481	0.001370544	8	9	9
TW_Artery_Aorta_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	1.201591	0.044835	0.229522	0.018118	0.021501	0.039769694	0.022096813	18	23	23
TW_Artery_Tibial_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	0.904067	0.04806	0.36596	0.007106	0.012703	0.057380821	0.0238277	4	6	6
TW_Brain_Putamen_basal_ganglia_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	1.426944	0.015032	0.153596	0.208912	0.082556	0.00886192	0.017125428	43	48	48
TW_Cells_EBV-transformed_lymphocytes_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	-0.76219	-0.01427	0.445946	0.054498	0.035753	0.04392222	0.041637927	21	26	26
TW_Esophagus_Mucosa_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	-0.07335	0.001549	0.941526	0.065061	0.145442	9.20E-10	1.43E-09	33	39	39
TW_Heart_Atrial_Appendage_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	-0.56168	-0.01964	0.574336	0.020932	0.025039	0.046356302	0.034806914	26	29	29
TW_Heart_Left_Ventricle_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	0.655161	0.011925	0.512364	0.075275	0.141574	8.79E-08	2.59E-07	20	23	23
TW_Lung_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	0.489939	0.039174	0.624177	0.024272	0.024632	0.008760454	0.005973537	20	21	21
TW_Muscle_Skeletal_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	1.749909	0.071652	0.080134	0.0118	0.104255	3.33E-10	5.83E-10	10	13	13
TW_Pancreas_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	-0.63832	-0.01156	0.523264	0.066281	0.056007	0.003662688	0.004143342	18	24	24
TW_Pituitary_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	0.650747	0.00577	0.51521	0.33837	0.194396	1.92E-05	0.000105887	67	78	78
TW_Skin_Not_Sun_Exposed_Suprapubic_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	1.130898	0.020226	0.258098	0.094501	0.188317	2.10E-10	8.25E-10	20	23	23
TW_Skin_Sun_Exposed_Lower_leg_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	1.713745	0.040153	0.086576	0.118149	0.171147	6.47E-14	1.40E-13	52	62	62
TW_Spleen_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	0.883423	0.015798	0.377008	0.059964	0.048955	0.037183669	0.036604122	26	28	28
TW_Stomach_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	1.429465	0.026123	0.152871	0.044824	0.053373	0.002436732	0.003789078	13	14	14
TW_Whole_Blood_0.5.db.CRP.csv:ENSG00000185033.10	Inflammation	SEMA4B	0.363388	0.007313	0.716315	0.088917	0.121142	4.64E-11	7.68E-11	30	33	33
TW_Adipose_Visceral_Omentum_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	0.068051	0.007986	0.945745	0.011719	0.058048	0.000954481	0.001370544	6	9	9
TW_Artery_Aorta_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-1.20747	-0.15445	0.22725	0.023666	0.021501	0.039769694	0.022096813	15	23	23
TW_Artery_Tibial_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-0.76085	-0.45061	0.446749	0.000618	0.012703	0.057380821	0.0238277	4	6	6
TW_Brain_Putamen_basal_ganglia_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-0.30872	-0.01316	0.757534	0.050781	0.082556	0.00886192	0.017125428	24	48	48
TW_Cells_EBV-transformed_lymphocytes_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-1.26112	-0.10196	0.207266	0.035338	0.035753	0.04392222	0.041637927	15	26	26
TW_Esophagus_Mucosa_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	0.708132	0.033001	0.478864	0.112392	0.145442	9.20E-10	1.43E-09	23	39	39

TW_Heart_Atrial_Appendage_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-0.538	-0.05781	0.590577	0.019937	0.025039	0.046356302	0.034806914	19	29	29
TW_Heart_Left_Ventricle_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-0.85409	-0.07551	0.393054	0.030514	0.141574	8.79E-08	2.59E-07	12	23	23
TW_Lung_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-0.25233	-0.03736	0.800788	0.018843	0.024632	0.008760454	0.005973537	10	21	21
TW_Muscle_Skeletal_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-0.65097	-0.07568	0.515066	0.017356	0.104255	3.33E-10	5.83E-10	7	13	13
TW_Pancreas_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	0.038939	0.028788	0.968939	0.024748	0.056007	0.003662688	0.004143342	13	24	24
TW_Pituitary_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-1.259	-0.03792	0.208032	0.271517	0.194396	1.92E-05	0.000105887	49	78	78
TW_Skin_Not_Sun_Exposed_Suprapubic_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	0.134409	-0.00241	0.893079	0.051952	0.188317	2.10E-10	8.25E-10	14	23	23
TW_Skin_Sun_Exposed_Lower_leg_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-1.64358	-0.09774	0.100264	0.070973	0.171147	6.47E-14	1.40E-13	33	62	62
TW_Spleen_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-1.8754	-0.13162	0.060738	0.045648	0.048955	0.037183669	0.036604122	16	28	28
TW_Stomach_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	-1.92961	-0.16994	0.053655	0.033186	0.053373	0.002436732	0.003789078	7	14	14
TW_Whole_Blood_0.5.db.PGC.MDD.csv:ENSG00000185033.10	Major depression	SEMA4B	1.294453	0.095496	0.195509	0.048565	0.121142	4.64E-11	7.68E-11	17	33	33

eTable 6: Significantly associated (*cis*) SNPs with the top CpG sites

PValue	SNPName	SNPChr	SNPChrPos	ProbeName	ProbeChr	ProbePos	CisTrans	SNPType	AlleleAssessed	OverallZScore	Total N	eQTMGene	FDR
6.72E-122	rs9880418	3	56809820	cg12325605	3	56810175	cis	A/G	A	-23.4786941	3841		0
7.41E-09	rs751837	14	103484825	cg04987734	14	103415849	cis	T/C	C	-5.781334	3841	CDC42BPB	6.09E-06
1.41E-08	rs3821412	3	56787451	cg12325605	3	56810175	cis	T/C	T	5.6722757	3841		1.47E-05
1.95E-06	rs13081650	3	56941112	cg12325605	3	56810175	cis	T/C	C	-4.7589465	3841		0.001692
1.18E-05	rs4398049	14	103438905	cg04987734	14	103415849	cis	A/G	A	-4.381021	3841	CDC42BPB	0.007461
0.0001065	rs78561280	3	56889061	cg12325605	3	56810175	cis	C/T	T	3.8752347	3841		0.048648

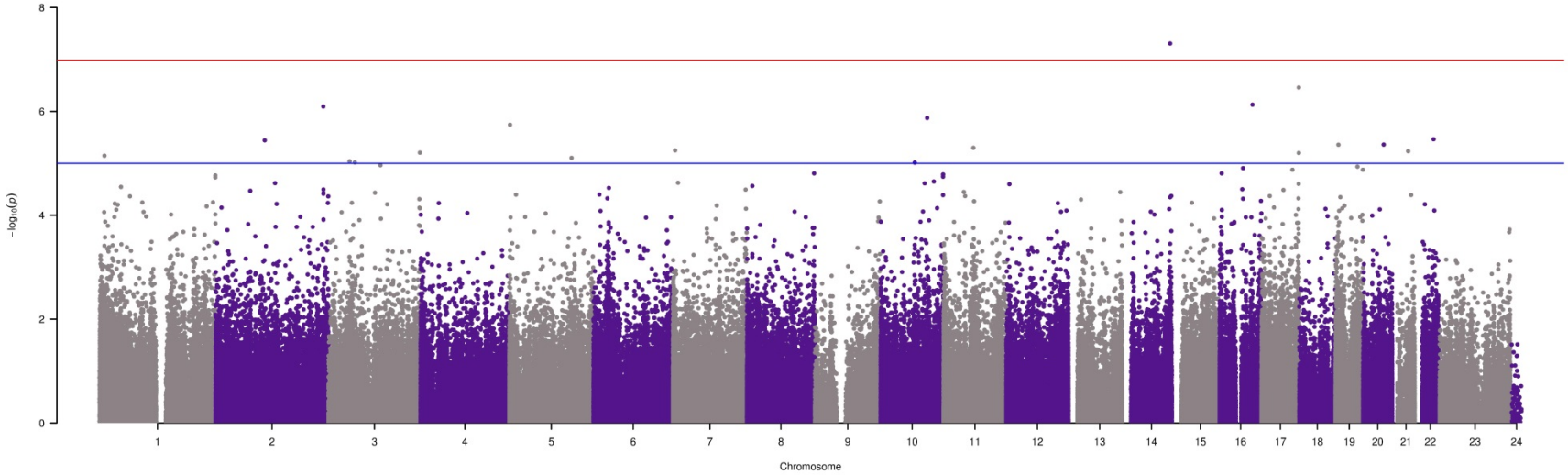
eTable 7: Results of causal inference analyses

CpG site		cg04987734			cg12325605	
cis-SNP		rs751837			rs3821412	
TRAITS	effect allele	Effect	Pvalue	effect allele	Effect	Pvalue
MDD	T	0.9415	0.0786	T	1.02	0.2906
CRP	T	0.0055	0.5807	T	-0.0048	0.442
CPD	T	0.2179	0.08687	T	0.039	0.6344

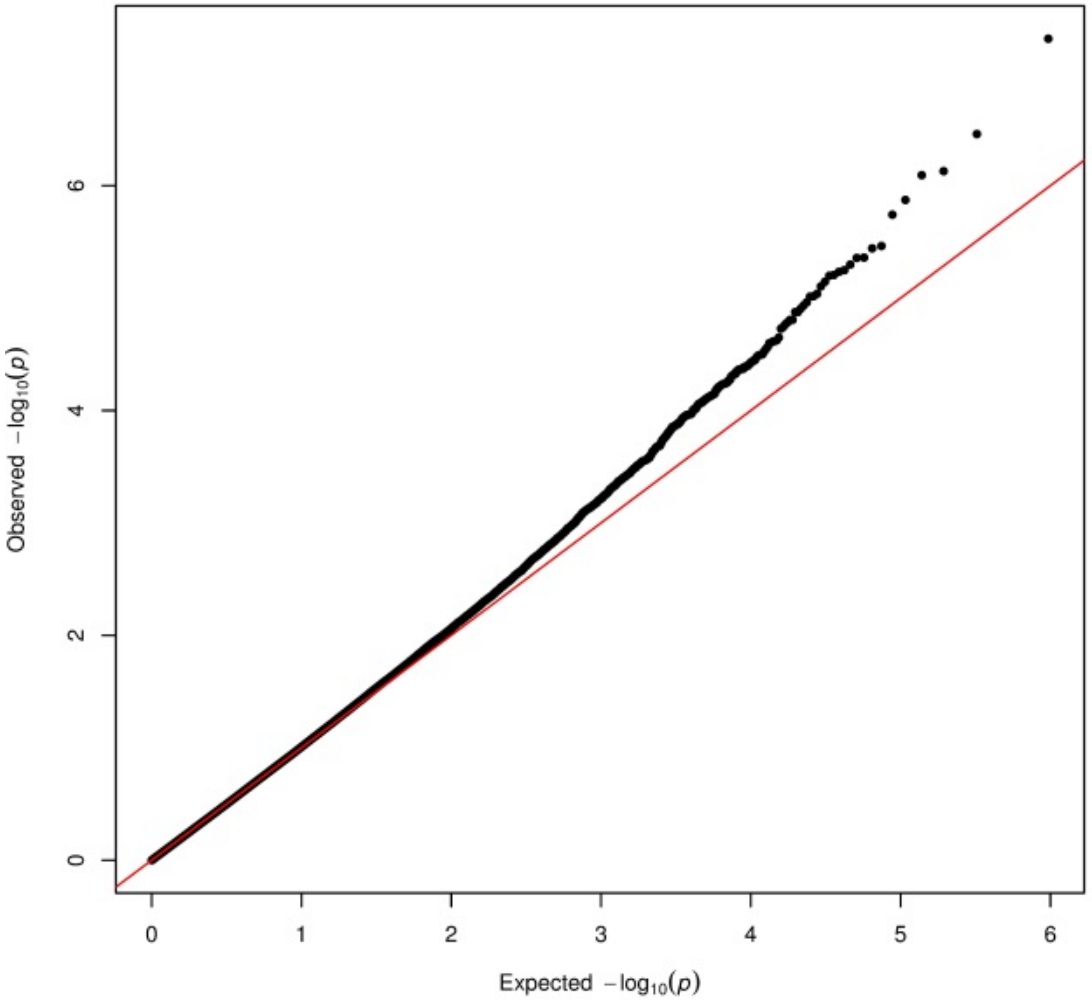
MDD: major depressive disorder
CRP: C-reactive protein
CPD: Cigarettes per day

eFigure 1: Genome-wide association plot of the discovery EWAS.

Horizontal axis depicts chromosomes and vertical axis depicts the negative logarithm of the association *p-value*. Each dot represents a CpG site. The solid red line indicates the genome-wide significance threshold and blue solid line indicates the suggestive threshold.

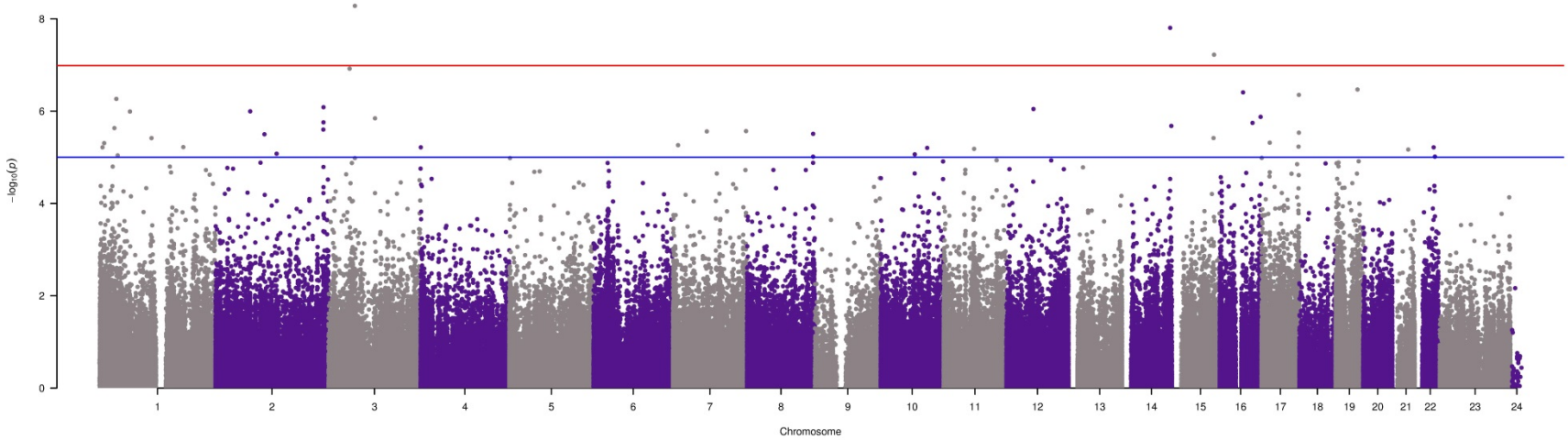


eFigure 2: Quantile-quantile plots for the discovery EWAS. Horizontal axis depicts the expected negative logarithm of the association *p-value*, while the vertical axis shows the observed negative logarithm of the association *p-value* (λ = 1.030).

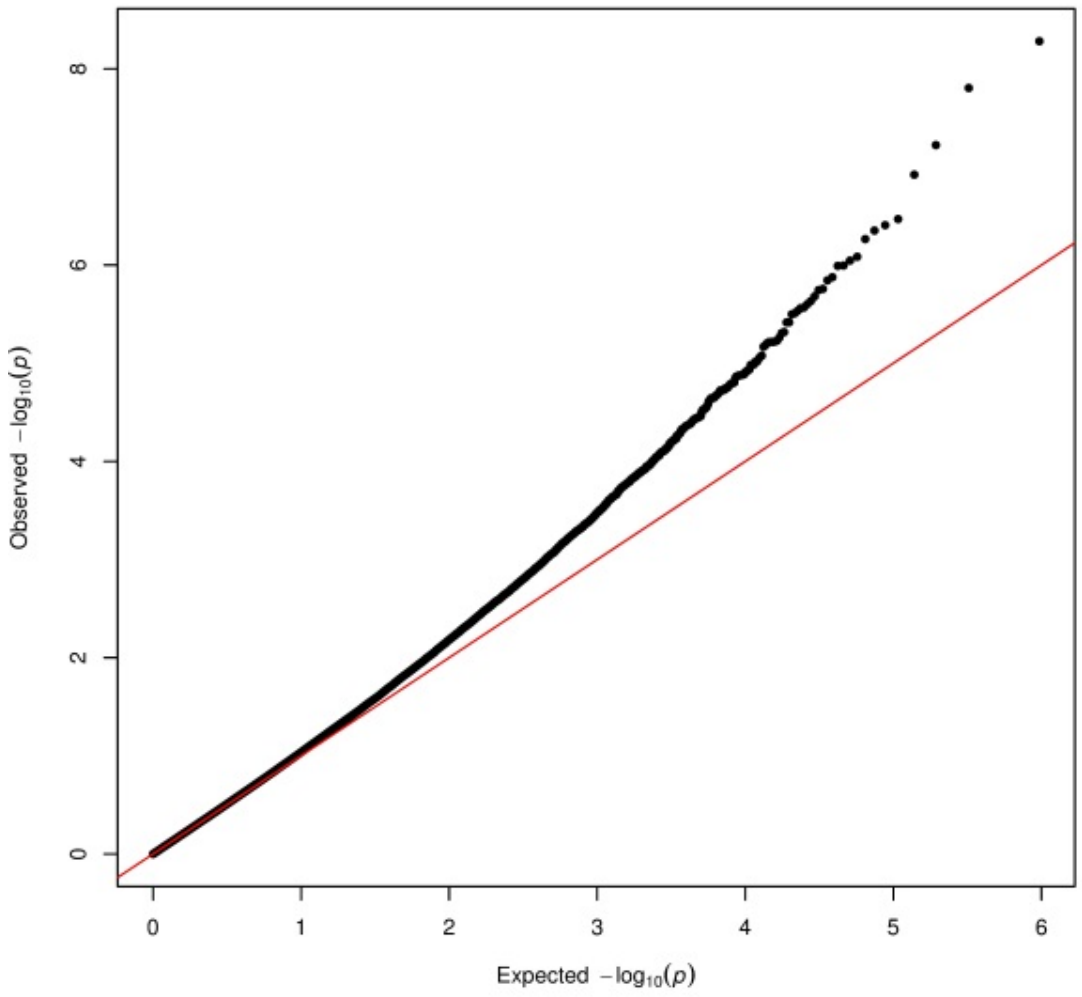


eFigure 3: Genome-wide association plot of the meta-analysis of discovery and replication EWAS.

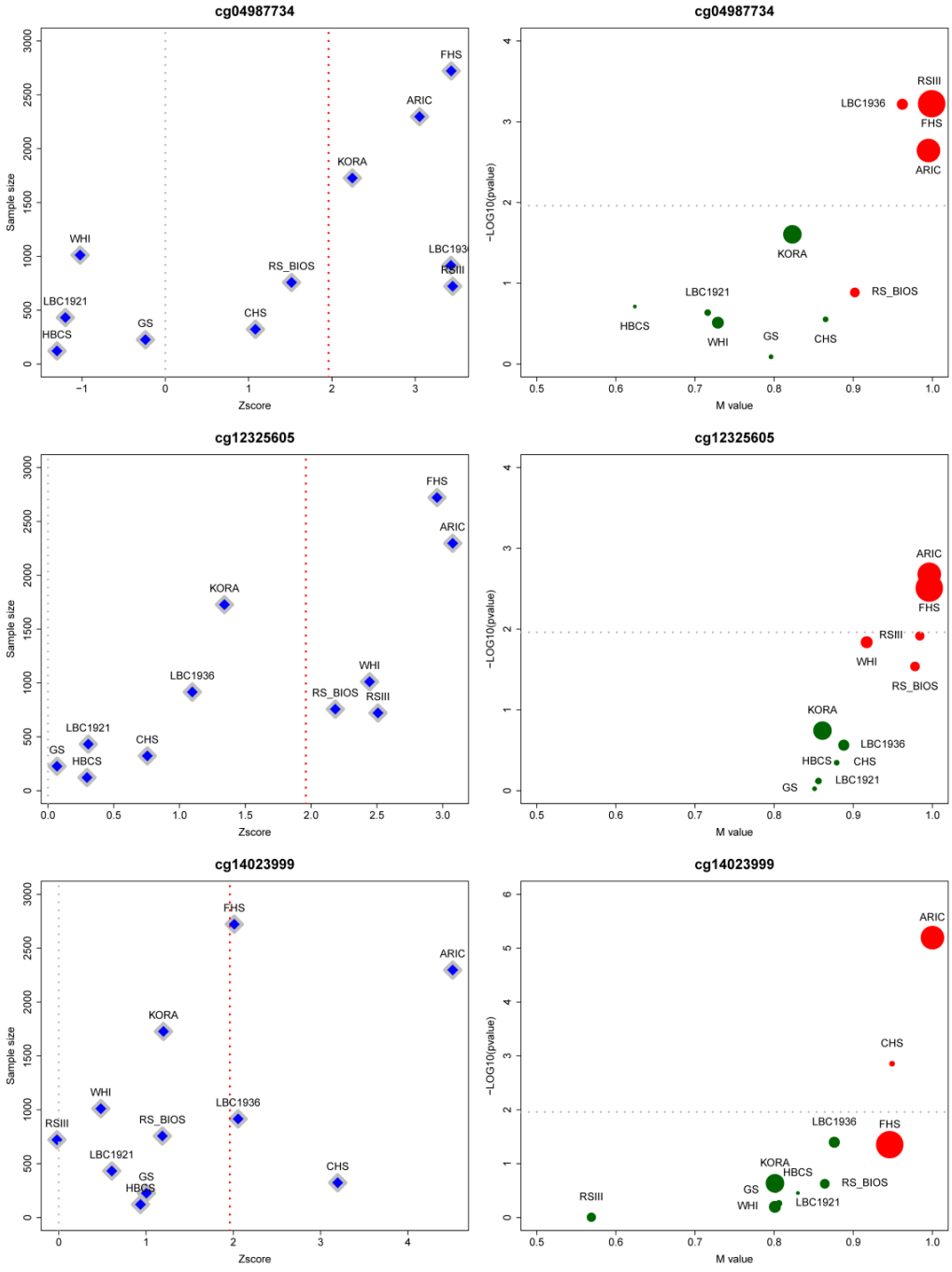
Horizontal axis depicts chromosomes and vertical axis depicts the negative logarithm of the association *p-value*. Each dot represents a CpG site. The solid red line indicates the genome-wide significance threshold and blue solid line indicates the suggestive threshold.



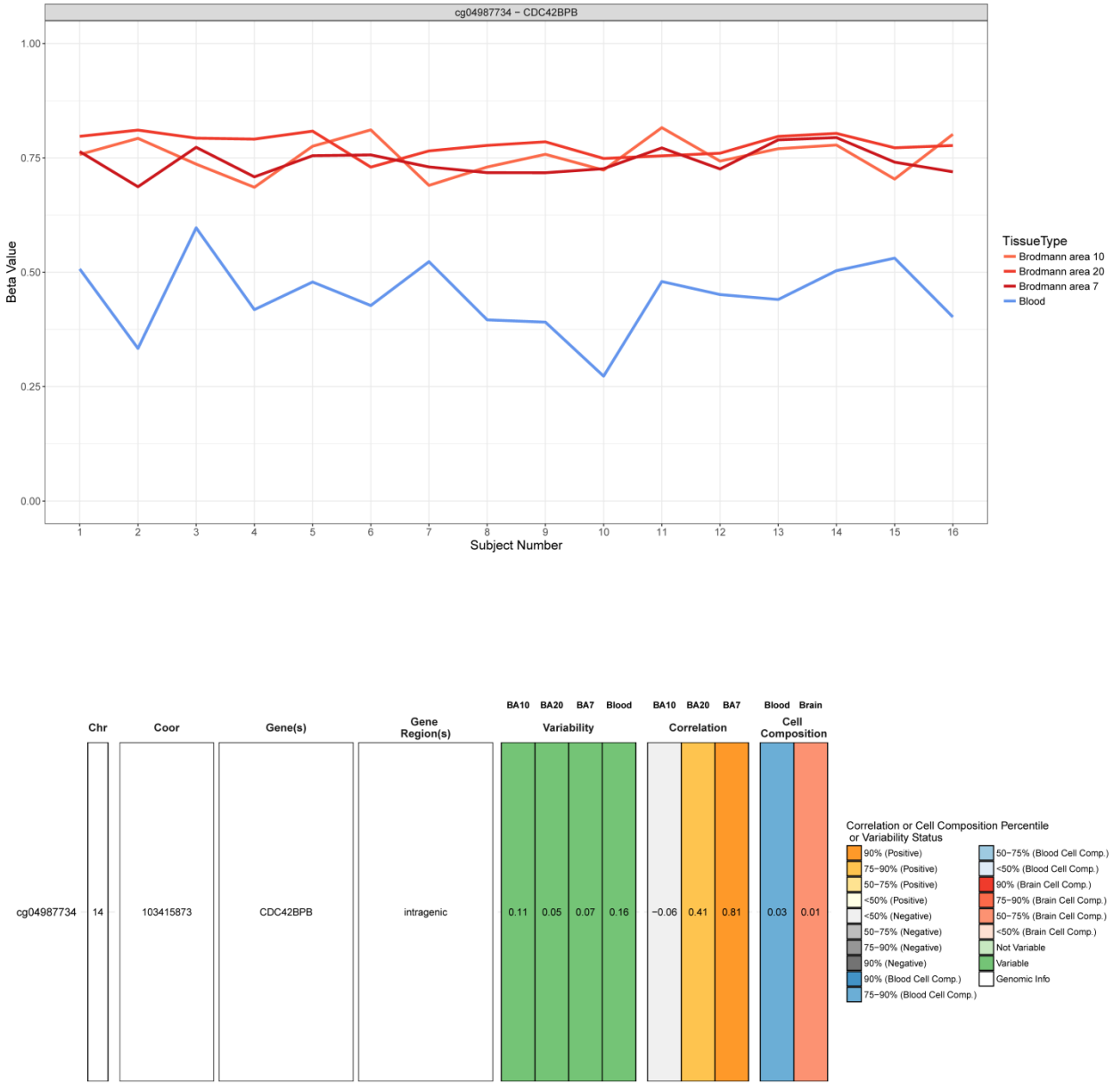
eFigure 4: Quantile-quantile plot of the meta-analysis of discovery and replication EWAS (λ = 1.088). Horizontal axis depicts the expected negative logarithm of the association *p-value*, while the vertical axis shows the observed negative logarithm of the association *p-value*.



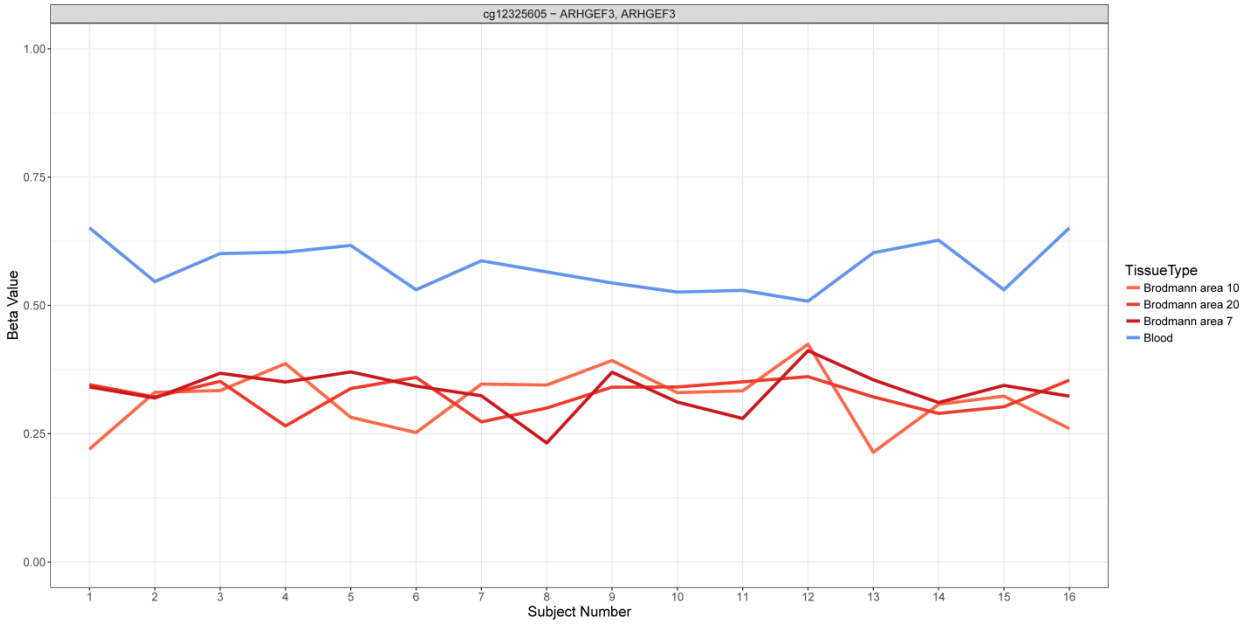
eFigure 5: Forest and M-value plot for the three significant CpG sites. The left panels illustrate the association z-scores for the CpG site for each study against the sample size of the study. The red dotted vertical bar represents the nominal significance level (z-score < 1.96). The right panels depict the posterior probability of effect in each study (M-value) against the negative logarithm of the p-value for that study. Red dots suggest that the study has an effect (m-value > 0.9), green dots suggest that the study's effect is uncertain (0.1 < m-value < 0.9), while blue dots (not observed in our case) would suggest that the study has no effect (m-value < 0.1).

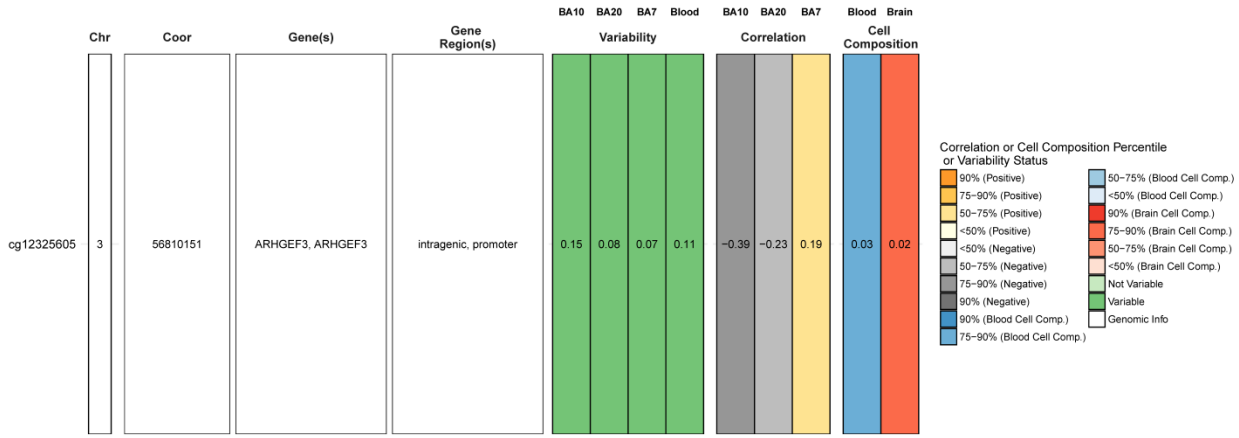


eFigure 6. Correlation between blood and brain methylation at cg04987734.

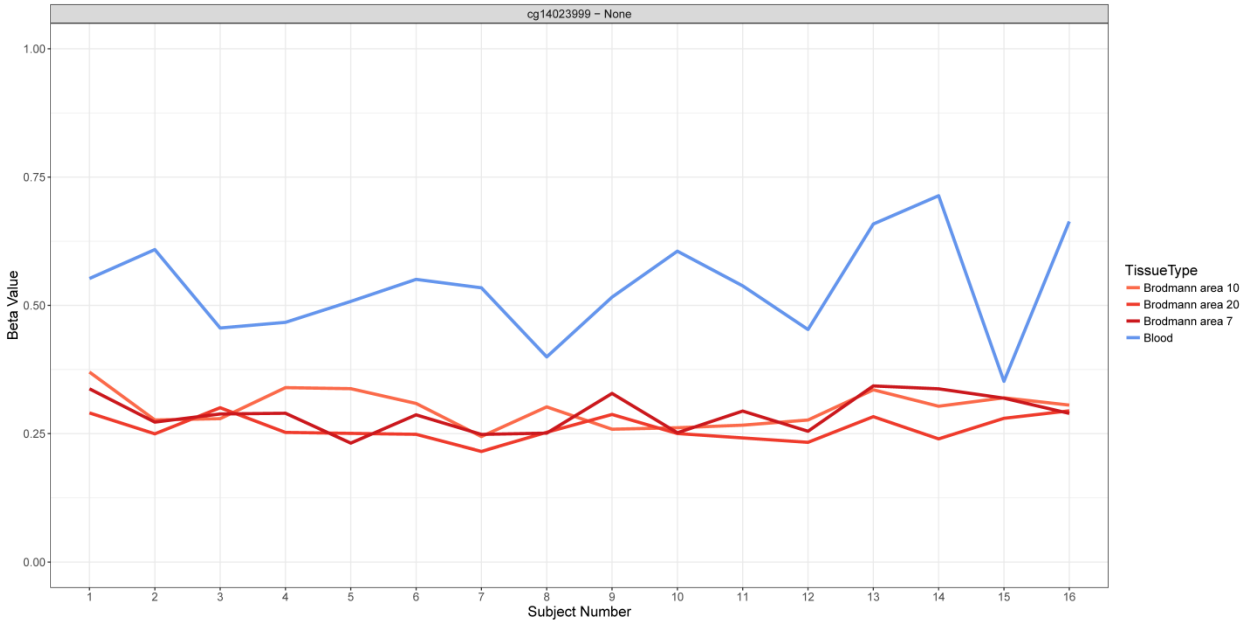


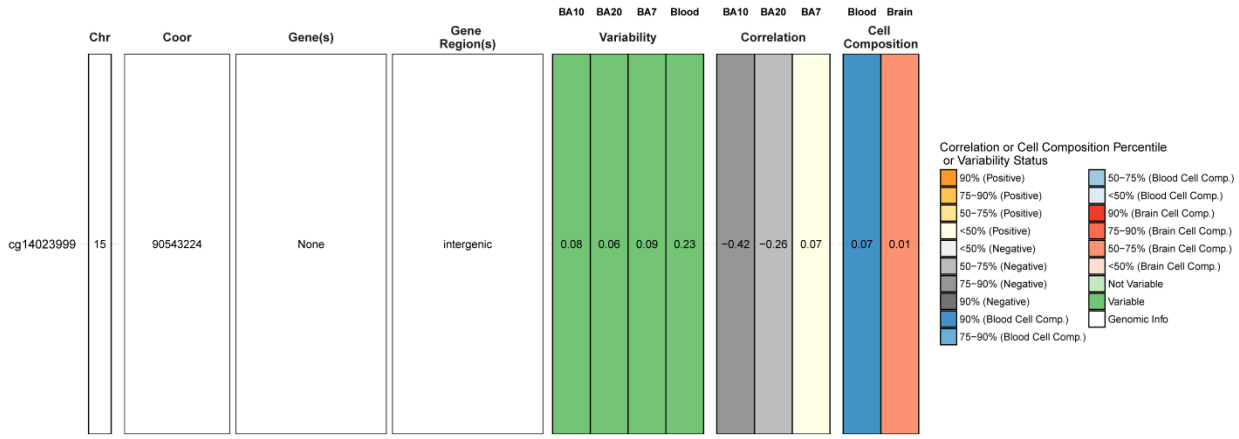
eFigure 7: Correlation between blood and brain methylation at cg12325605.



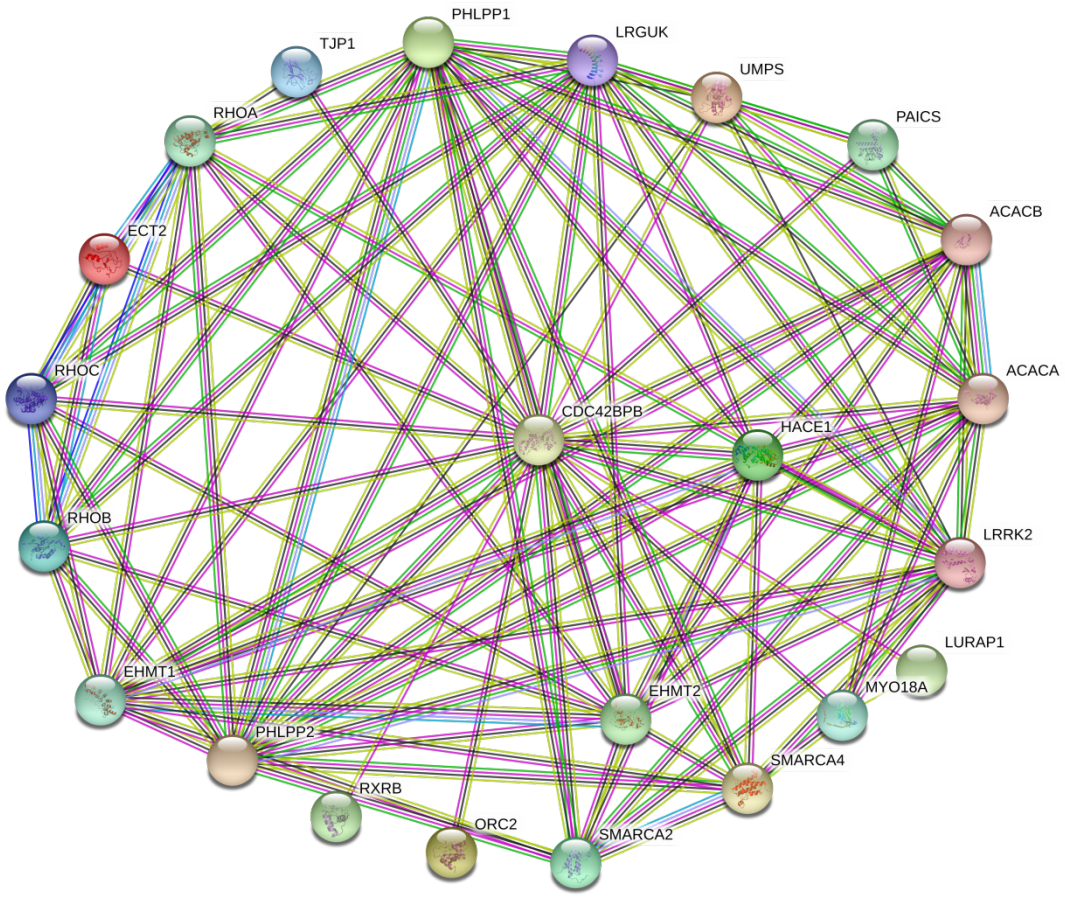


eFigure 8: Correlation between blood and brain methylation at cg14023999.





eFigure 9: Gene network of *CDC42BPB* using STRINGS network database



eFigure 10: Gene network of *ARHGEF3* using STRINGS network database